NEW USES OF SECOND ORDER PROBABILITY TECHNIQUES IN ESTIMATING CRITICAL PROBABILITIES IN COMMAND & CONTROL

D. Bamber* & I.R. Goodman *
Code D44215, Space & Naval Warfare Systems Center,
San Diego, CA 92152
Telephone Numbers: 619-553-4014
Email Addresses:bamber@spawar.navy.mil, goodman@spawar.navy.mil

Abstract

It is an understatement that both the theory and applications of probability –conditional or unconditional – play an essential role in the processing and use of disparate information in decision-making in C4I systems. Apropos to the theme of this symposium, “Making Information Superiority Happen”, the paper outlined here describes new applications, insights, and theoretical aspects of ongoing work by the authors toward improving the rationale for use of probability theory, keeping in mind issues of scalability and computational complexity. This paper extends the ideas first presented in last year’s CCRTS at Newport, RI. In short, the mathematical theme of this paper is both a summary of past research efforts together with new results on the problem of best estimating partially specified conditional and unconditional probabilities of interest via a second order bayesian probability approach. Among the new derivations provided in this paper is a significant reduction in computational effort in obtaining (again, in the second order probability sense) optimal or “near-optimal” probability estimates, all within the setting of a boolean “conditional event algebra” which allows full compatibility with conditional probability evaluations.

1. Introduction

As stated in the abstract, this work is a direct continuation of the effort presented in [Goodman, 1999]. Even in the simplest appearing situation, where probabilistic information is present in the form of specified or apriori known (or estimated) probabilities of certain contributing events, the theory of how to determine or best estimate the probability of another particular event, or events of interest, may not be readily apparent. In addition, it is possible that this problem cannot be resolved within the confines of ordinary probability theory because at times it seems to be at odds with our “commonsense” solution. This phenomenon is seen to occur even at the simplest levels, as will be illustrated later. Since a basic aspect of reasoning relative to Command & Control relies heavily upon probability concepts, these issues must be resolved within a framework of rigor, yet computational tractability. Such real-world probabilities are often not necessarily even fully theoretically determined. This typically occurs in rule-based systems where all that is known concerning conditional probabilities associated with rules are lower bounds on those probabilities. In addition, many popular techniques, such as Bayes nets (see,

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **2000**	2. REPORT TYPE	3. DATES COVERED **00-00-2000 to 00-00-2000**

4. TITLE AND SUBTITLE **New Uses of Second Order Probability Techniques in Estimating Critical Probabilities in Command & Control**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center San Diego,53560 Hull Street,San Diego,CA,92152-5001**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
2000 Command and Control Research and Technology Symposium

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**53**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

* This joint work was supported by the FY00 Office of Naval Research (ONR) In-House Laboratory Independent Research Program at the Space & Naval Warfare Systems Center, San Diego, Program ZU58.

e.g., [Pearl, 1988] for a basic exposition), may not be applicable, unless one is willing to make a number of independence assumptions – assumptions for which there may not always be good justification. This paper will address a number of issues connected with the estimation of underspecified probabilities.

1.1 *Notation, Conventions, Fundamental Results*

Recalling the basic identification between classical propositional logic (actually also including its extension to quantified logic) and boolean (and sigma) algebra of events or sets, as given through the Stone representation theorem provided in details, e.g., in Chapter 5 of [Mendelson, 1970], we choose throughout this paper to present all concepts via a boolean algebra framework and its ramifications. Here, letters a, b, c,... indicate events (or statements which can be true or false) in a boolean algebra *B*, where: all disjunctions or unions of events in *B* are indicated by $\vee$ (where any finite quantity of such disjunctions produces another event in *B*); all complements or negations of events by $(.)'$, as, e.g., a' (also an event in *B*); and all conjunctions or intersections of events by either &, as e.g., a&b, or when possible, by simply omitting any symbol such as ab. Repeated conjunction or disjunction operations are indicated in the usual way in string form or with use of an index set below the symbols & or $\vee$. At times, capital letters in roman form A, B,..., or in italic form *A*, *B*,..., as well as lower case greek letters α, β, γ,..., will be used to indicate special events or special collections of events. The universal event or set containing all events of relevance to the problem at hand is indicated by Ω, while the null event or set is indicated by $\emptyset$. The standard subevent partial order relation is denoted as $c \leq d$. Equality of events is simply denoted by =, etc. The triple (Ω,B,P) refers to a probability space with probability measure $P:B\to[0,1]$ (unit interval). Here, generally, *B* is a boolean algebra, but if needed, *B* may also be a sigma algebra (where all countable infinite repetitions of & and $\vee$ on events in *B* produce again events in *B*). When necessary to distinguish between the boolean algebra operators acting on events and logical operators acting upon sets of events or upon index sets relative to the events, we use the set notation $\cap$, $\cup$, $\subseteq$, rather than the corresponding &, $\vee$, $\leq$, etc. When needed to emphasize a point, we use the convention $=^{d}$ to indicate a definition, rather than a proved result and $=^{w}$ to indicate "which equals". The notation card(J) refers to the cardinality of (usually, index) set J.

Metalogical notation – i.e., notation utilized in proving theorems or making remarks *about* boolean algebras, probabilities, etc. involved in the results -- will employ ordinary "and", "or", "not", "if-then" or "implies", "iff" for "if and only if", i.e., logical equivalence, etc.

Multivariable notation will be applied when more efficient than writing out arguments or vector components. For example the family of events $(a_j)_{j \text{ in } J}$ for some finite index set J, can also be denoted as simply a_J, the repeated conjunction $\&(a_j)$ can also be denoted as $\&(a_J)$; $\&(a'b)_J$ for $\underset{j \text{ in } J}{\&}(a_j{}'b_j)$; $\vee(a_J)$ for $\underset{j \text{ in } J}{\vee}(a_j)$; $P(a_J) = 0_J$, for $P(a_j) = 0$, j in J. Also, 1_m is that m by 1 vector, each of whose one-dimensional components is 1, with an analogous definition for 0_m. More generally, when unambiguous, 1_J is that vector of card(J) components, each being 1, etc. For any matrix or vector A, sum(A) is simply the sum of all of its elements. In a related vein, iterated summations

such as $\sum_{j \text{ in } J}(P(a_j))$, $\sum_{j \text{ in } J}(P(a_j \mid b_j))$, $\sum_{j \text{ in } J}(P(a_j b_j))$ all will be streamlined, whenever unambiguous, to be, respectively, $\Sigma(P(a)_J)$, $\Sigma(P(a|b)_J)$, $\Sigma(P(ab)_J)$, etc. Further multivariable notation will be provided as needed. In addition, four special binary boolean operators at times will be of use and are indicated in action for any a, b in *B* as:

(i) *material conditional or logical implication* – the classic logic truth-table counterpart (or event indicator function) is only false when the antecedent is true and the consequent is false, and in a sense, most naturally models "if-then" or conditional statements from a classical logic viewpoint

$$\text{"if b, then a" becomes } b \Rightarrow a =^d b' \vee a =^w b' \vee ab =^w (a'b)' =^w (a \neg b)', \tag{1.1.1}$$

where "¬" is defined next.

(ii) (non-symmetric) *event difference*

$$\text{"b and not(a)" becomes } b \neg a =^d a'b. \tag{1.1.2}$$

(iii) *symmetric event difference*

$$\text{"(b and not(a)) and (a and not(b))" becomes } a \Delta b =^d (a'b) \vee (b'a) =^w (a \Leftrightarrow b)', \tag{1.1.3}$$

where ⇔ is defined next.

(iv) *logical equivalence*

"a iff b" or "(if b, then a) and (if a, then b)" becomes

$$a \Leftrightarrow b =^d (ab \vee a'b') =^w ((b \Rightarrow a)\&(a \Rightarrow b)) =^w (a \Delta b)'. \tag{1.1.4}$$

In addition to being aware of the elementary properties of probabilities, we will need on several occasion to make use of the Fréchet-Hailperin-Hoeffding tightest general probability bounds on conjunction (see, e.g., [Hailperin, 1965, 1984]): For any given probability space (Ω,*B*,P) and events a_j in *B*, j in J, for some finite index set J,

$$\max\Big(\sum_{j \text{ in } J}(P(a_j)) - (\text{card}(J) - 1), 0\Big) =^d L(a_J,P) \leq P(\&(a_J)) \leq U(a_J,P) =^d \min_{j \text{ in } J}(P(a_j)). \tag{1.1.5}$$

From now on, the inequalities in eq.(1.1.5) will be referred to as the *FHH inequalities*.

Finally, a word on when conditional probabilities are well defined. Generally speaking, conditional probabilities such as P(a|b) are only meaningful when P(b) > 0, in which case the standard definition holds

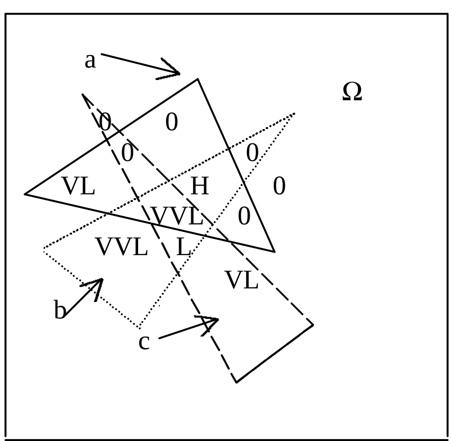
$$P(a|b) =^d P(ab)/P(b). \tag{1.1.6}$$

But, at times, individuals have found it convenient to extend the standard definition to either yielding unity or zero for the case when P(b) = 0. In particular, Adams has made such use in developing his probability logics [Adams, 1996]. While later we will consider the situation where in fact P(b) = 0 (although not actually defining P(a|b) directly for that case), we will distinguish carefully through this paper between the standard case – when P(b) > 0 – and the nonstandard. *From, now on, whenever the simple symbol P(a|b) is used, it is assumed that P(b) > 0.*

1.2 *One Motivation for the Work: The Transitivity Problem*

To illustrate the, unfortunately, all-to-often occurring discrepancy between probabilistic and commonsense reassuming, consider the following example. First, let us abbreviate the following statements/events: b $=^d$ "enemy is secretly amassing over 100,000 troops ready to attack"; c $=^d$ "political negotiations will fall through and it will be foggy tomorrow morning"; and a $=^d$ "enemy will attack tomorrow morning". Suppose in this situation that previously acquired intelligence information indicates that all three events are neither certain nor impossible and that estimates of the following two conditional probabilities are the *only* reliable available information: P(a|b) = 0.9 (approximately) and P(b|c) = 0.8 (approximately).

What can we say about the critical desired probabilities such as P(a) or P(a|c) ? Using the basic laws of probability, it can be shown that, unless we make further assumptions, the above two probabilities can take essentially any values in the unit interval. More specifically, Figure 1 illustrates why, in general, with lack of any specific assumptions, one could have both P(a|b) and P(b|c) very high, but P(a|c) low or even zero. There the triangles represent any three overlapping events a, b, c and the enclosing rectangle represents Ω. A probability measure P is chosen with mass to be distributed over a, b, c so that, as usual P(Ω) = 1. The probability assignments are shown for the various regions (or relative atoms) scoped out by conjunctions of combinations of affirmations and negations of a, b, c, where VVL indicates "very, very low" (but not zero), VL indicates "very low", L indicates "low", and H indicates "high".



P(abc) = VVL, P(abc′) = H,
P(a′bc) = L, P(a′b′c) = VL,
P(ab′c′) = VL, P(a′bc′) = VVL.
(Typically here, VVL = 0.001, VL = 0.01,
L = 0.1, H = 0.878.)

$P(a|b) = (H + VVL) / (H + L + 2VVL) \approx 1$
(Typ. Val: 0.897)

$P(b|c) = (L + VVL) / (L + VL + VVL) \approx 1$
(Typ. Val: 0.910)

$P(a|c) = (VVL) / (L + VL + VVL) \approx 0$
(Typ. Val: 0.0090)

Figure 1. Example for conditional probability extension of classical transitivity-syllogism problem where premise set has high conditional probability values, but conclusion has a low (or even zero) conditional

probability value.

Let us return to the general problem of given knowledge of $P(a|b)$ and $P(b|c)$, to determine $P(a|c)$ in some way. If we knew also $P(a|\ bc)$ and $P(a|\ b'c)$, then using also $P(b|c)$ (without using $P(a|b)$), we could easily determine $P(a|c)$, as the reader can check. Or, if we make the P-conditional independence assumption that b alone without c is "sufficient" for a, i.e., $P(a|\ bc) = P(a|b)$, and again, that $P(a|\ b'c)$ is now known, we could determine $P(a|c)$ completely by elementary probability considerations, using both $P(a|b)$ and $P(b|c)$, as in

$$P(a|c) = P(ab \mid c) + P(ab' \mid c) = P(a \mid bc)P(b|c) + P(a|b'c)P(b'|c) = P(a|c)P(b|c) + P(a|b'c)(1-P(b|c)). \quad (1.2.1)$$

Still other assumptions can be made about a, b, c and P to estimate $P(a|c)$ in some sense. (See, e.g., Section 6 of [Bamber *et al.*, 2000].)

On the other hand, *intuitively speaking*, when $P(a|b)$ and $P(b|c)$ are both high, even though it is possible that some probability configuration between a, b, and c may even yield $P(a|c) = 0$, it seems that *on the average* (whatever that means !) $P(a|c)$ should also be high – though possibly somewhat lower than both 0.9 and 0.8. This intuitively desirable property, called at times *transitivity*, *chaining*, or *hypothetical syllogism*, in the literature concerning the extension of classical reasoning to a probability framework [Bamber et al., 2000; Pearl, 1988] itself, has been the center of much controversy over the past several years in attempting to design rule-based systems which follow the laws of probability, but also agree with commonsense reasoning as the above example illustrates. One basic reason for this is that rule-based systems usually operate upon the sequential "firing" of rules, i.e., when the antecedent of one rule matches the consequent of another. But -- as is often the case -- when such rules are actually not 100% reliable, but for purpose of convenience (and the usual real world tradeoff in using something that is highly reliable but not perfect) still form part of the system, in effect, the transitivity problem is present, even if one tacitly ignores it to perform the functioning of the system. In a related vein, one should mention tacit alternatives to the problem of extending transitivity and other desirable properties of reasoning systems to a probability framework via "certainty factors" utilized in [Buchanan & Shortliffe, 1984] and other ad hoc procedures for combining reliabilities of inference rules, as discussed in [Hayes-Roth *et al.*, 1983].

Note also that, instead of interpreting the above conditional statements via naturally corresponding conditional probabilities, the statements "if b, then a", "if c, then b", and "if c, then a", could be first modeled through the classical logic (or boolean algebra) material conditional operator and then evaluated probabilistically. In that case, it easily follows that since

$$(b\Rightarrow a) \,\&\, (c\Rightarrow b) = b'c' \vee ab \leq c' \vee a = c \Rightarrow a, \quad (1.2.2)$$

by the monotonicity property of probability, for any P over *B*,

$$P((b\Rightarrow a) \,\&\, (c\Rightarrow b)) \leq P(c \Rightarrow a), \quad (1.2.3)$$

and applying the lower bound FHH to eq.(1.2.3), we obtain

$$[P(b\Rightarrow a) \geq s,\ P(c\Rightarrow b) \geq t\ ] \text{ implies } [P(c \Rightarrow a) \geq s + t - 1], \quad \text{for any } ½ \leq s, t \leq 1, \tag{1.2.4}$$

and thus have a seemingly satisfactory solution to the transitivity problem (where both the probabilistic analysis and commonsense reasoning apparently agree).

However, one basic fact that precludes a consistent use of $(b\Rightarrow a)$ in interpreting the conditional expression "if b, then a", and the subsequent probability evaluation $P(b\Rightarrow a)$ for measuring the degree of reliability of the rule "if b, then a", or the uncertainty of same rule, is that the evaluation

$$P(b\Rightarrow a) = 1- P(b) + P(ab) \tag{1.2.5}$$

increases toward unity when $P(b)$ decreases down toward zero, regardless of the relationship between $P(b)$ and $P(ab)$ – which, of course $P(a|b)$ completely respects. Moreover, a well-known inequality provides a good quantitative measure of the difference between the two approaches to modeling (see, e.g., [Calabrese, 1987; Goodman & Nguyen, 1995] for further discussions)

$$P(b\Rightarrow a) = P(a|b) + P(b')P(a'|b) \geq P(a|b). \tag{1.2.6}$$

On the other hand, when $P(a|b) = 1$, $P(b|c) = 1$, a little manipulation shows necessarily $P(a|c) = 1$. In fact, this case generalizes the classical (*barbara*-type) of syllogism typified by the well-known paradigm "All men are mortal", "I am a man", therefore "I am mortal" (See, e.g., [Prior *et al.*, 1967; Copi, 1986; Goodman, 1999] for discussion and background on this classical syllogism.) Thus, we see that a real sort of discontinuity exists between the certain conditional probability or classical reasoning case and the general nontrivial conditional probability case for potential transitivity, keeping in mind the additional difficulty illustrated above that the material conditional-plus-probability-evaluation approach is also not satisfactory, despite its formally satisfying transitivity at all probability levels.

Besides the transitivity problem, a number of other fundamental problems exist in reasoning which also yield similar apparent discrepancies with our commonsense understanding, including *contraposition* and *strengthening*, discussed later. In the case of all of the above-mentioned problems, the desired probability subject to the given constraints is so underspecified that in general it can range over the entire unit interval. In such situations it appears that a number of previously established approaches to estimating varying probabilities may not be adequate. This includes the many general bounding, probability-bounding, upper and lower probability techniques, and random set and related (belief, plausibility, etc.) function techniques, as provided, e.g., in [Alefeld & Herzberger, 1983; Hailperin, 1996; Walley, 1991; Dempster, 1967; Shafer, 1976; Goodman et al., 1997]. On the other hand, these techniques, used with appropriate caution (see, e.g., [Nguyen, 1978; Chapters 3, 4 of Goodman & Nguyen, 1985]) may provide a viable alternative to that which we present here in the subsequent sections. In yet another direction, there is the basic – or naive – maxumum entropy approach, which picks a specific P – and then uses that P to evaluate the desired conclusion probability -- through maximizing entropy, subject to the constraints of the problem. This, indeed, may also furnish a possible reasonable approach to these issues, as developed, e.g., in [Rödder, 2000], based on general principles as found, e.g., in [Kapur, 1994]. However, as in the bounding approaches, such use of maximum entropy must be carried out with caution, as will be seen later.

1.3 *Use of Second Order Probability in Addressing the Transitivity Problem*

In [Goodman & Nguyen, 1998, Goodman, 1999 and Bamber *et al.*, 2000] it was shown that a reasonable way to analyze the transitivity problem within a completely rigorous mathematical framework, compatible with all the laws of probability, is to interpret the above expression *on average* to mean that, instead of attempting hopelessly to pick out the "true" value P(a|c) from the unit interval, one should average the value P(a|c) *over all possible choices of probability measures* P, subject to the given constraints P(a|b) = 0.9 and P(b|c) = 0.8. But, this *bayesian method* requires a choice of *second order probability*, i.e., a choice of probability distribution of the probability measures themselves! (Second order probability techniques have already proven useful in addressing update problems as in [Goodman & Nguyen, 1999a] and may also be found in the older treatise of [Aitchison, 1986].) Suppose, for simplicity and lack of any other information, appealing, e.g., to a *second order* maximal entropy (or equivalently most ignorance of information argument) – as opposed to the naive (first order) maximum entropy approach discussed earlier -- we choose this second order distribution to be in a natural sense uniform over the possible candidate probability measures. Then, it can be shown [Goodman, 1999; Bamber *et al.*, 2000] that no matter what threshold values s = P(a|b) and t = P(b|c) are, a closed-form expression in variables s and t can actually be computed for the P-averaged P(a|c), which, in agreement with commonsense, does, in fact, approach unity as s and t approach unity. In addition, a reasonable upper bound can also be obtained for the error variance (between this estimate and actual possible values), as the probabilities vary uniformly. Computations for related procedures produce closed-form results in a number of cases of interest besides transitivity, but these approaches do not, at first analysis, appear to be generalizable, because of the difficulty in evaluating multiple integrals over spaces of constrained probability measures. However, recent efforts have produced promising modifications and approximations applicable to the general case, as outlined in this paper.

Returning to the case of transitivity, the formula for the averaged value of P(a|c) with respect to P varying uniformly and P(a|b) and P(b|c) known, as well as bounds on its variance and expected deviation from its limiting unity value, and bounds on the associated (second order) probabilities are given below:

$$E_P\big(P(a|c)\ \big|\ P(a|b) = s,\ P(c|d) = t\big) = st + f(s,t)\ , \quad 1 > s, t > ½, \tag{1.3.1}$$

where the correction term f(s,t) > 0 in eq.(1.3.1) is

$$f(s,t) =^{d} \ [(1-t)/2 \ - ([s(1-s)(2s-1)t(1-t^2)] / [t+2t^2 + (s(1-s)(1-t)(2+3t-t^2))])] > 0,\ \ 1 > s, t > ½. \tag{1.3.2}$$

In other words, for this generic example, the averaged-out value of P(a|c) is approximately the same as if the "conditional events "a given b" and "b given c" were P-independent, for all P, up to the correction term f. For the specific example at hand, where s = 0.9 and t = 0.8, the averaged value of P(a|c) is approximately 0.75 (as compared to 0.72 for the formal assumed independence). Note, as stated above, that as s, t approach unity, f(s,t) approaches zero and the

expectation approaches unity, agreeing with the commonsense argument. In mathematical notation this is

$$\lim_{(s,t\to 1)} \left[E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\right] = 1. \tag{1.3.3}$$

In turn, because all probabilities lie in the unit interval, the conditional variance

$$\begin{aligned}
&\mathrm{Var}_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\\
=\ & E_P\Big(\big[P(a|c) - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2 \mid P(a|b) = s, P(c|d) = t\Big)\\
=\ & E_P\big((P(a|c))^2 \mid P(a|b) = s, P(c|d) = t\big) - \big[E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2\\
\le\ & E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big) - \big[E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2\\
=\ & E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\cdot\big(1 - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big)\\
\le\ & 1 - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big),
\end{aligned} \tag{1.3.4}$$

where for values of $E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)$ not that close to unity, the second to the bottom expression in eq.(1.3.4) can be used as an upper bound estimate of the conditional variance. Thus, also the averaged deviation of P(a|c) from unity, which is

$$\begin{aligned}
&E\big((P(a|c) - 1)^2 \mid P(a|b) = s, P(c|d) = t\big)\\
&= E\Big(\big[P(a|c) - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2 + \big[E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big) - 1\big]^2 \mid\\
&\qquad P(a|b) = s, P(c|d) = t\Big)\\
&= \mathrm{Var}_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big) + \big[1 - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2,
\end{aligned} \tag{1.3.5}$$

has also a computable upper bound obtainable from use of eq.(1.3.5). The unity limiting form in eq.(1.3.3) shows that both the variance and averaged deviation of P(a|c) from unity both approach zero. Since the usual triangle inequality provides

$$\begin{aligned}
0 \le 1-P(a|c) \ \le\ & 1- E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\\
&+ \big| P(a|c) - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big|
\end{aligned} \tag{1.3.6}$$

the standard Chebychev inequality (see, e.g., page 95 of [Rao, 1973]) shows, for any $\lambda > 0$, denoting Prob as the corresponding second order (posterior) probability measure determined through the uniform prior we chose for the P's,

$$\begin{aligned}
&\mathrm{Prob}\Big(\big[P(a|c) - E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]^2 \ge \lambda \ \Big|\ P(a|b) = s, P(c|d) = t\Big)\\
&\le (1/\lambda)\cdot \mathrm{Var}_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big).
\end{aligned} \tag{1.3.7}$$

Combining eqs.(1.3.4), (1.3.6), (1.3.7), for any $\lambda > 0$,

$$\begin{aligned}
&\mathrm{Prob}\Big([1-P(a|c)]^2 \ge \lambda \ \Big|\ P(a|b) = s, P(c|d) = t\Big)\\
&\le \big(1- E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big) / \big(\lambda^{1/2} - \big[1- E_P\big(P(a|c) \mid P(a|b) = s, P(c|d) = t\big)\big]\big)^2.
\end{aligned} \tag{1.3.8}$$

Finally, eq.(1.3.3) applied to eq.(1.3.8) shows that

$$\underset{(s,t\to 1)}{\text{limit}} \Big(\text{Prob}\big([1\text{-}P(a|c)]^2 \geq \lambda \;\big|\; P(a|b) = s, P(c|d) = t\big)\Big) = 0, \qquad (1.3.9)$$

at least the rate described in eq.(1.3.8), so that in standard probability parlance, P(a|c) converges in (second order) probability to unity, under the conditions P(a|b), P(b|c) themselves converge to unity, assuming otherwise, any P is equally likely.

2. Use of Second Order Probability in Analyzing Whether Various Types of Desirable Reasoning Properties Can Be Extended to a Probability Setting: An Introduction

2.1 *Some Particular Examples*

As in the transitivity problem discussed in Section 1.3, a second order probability approach can be used to determine whether particular types of desirable properties that classical logic possesses -- or more general reasoning systems should possess -- carries over to a probability-based reasoning system.

In addition to *transitivity*, closed-form results have been obtained involving a number of other reasoning system properties, including *contraposition*, *positive conjunction*, and *strengthening*.

In the case of contraposition, the premise involved in natural language form is "if b, then a" and the (potential) conclusion is "if not(a), then not(b)", a well known valid classical logic property -- in fact, one will recall in boolean form this is the same as the identity

$$b \Rightarrow a = a' \Rightarrow b'. \qquad (2.1.1)$$

When the above forms are not only interpreted through the material conditional operator, but also evaluated via probability, just as in the attempt for addressing the transitivity problem (see eq.(1.2.4)), the classical logic property seems to extend trivially to a probability setting:

$$[P(b \Rightarrow a) \geq s\,] \;\text{ implies }\; [P(a' \Rightarrow b') \geq s], \quad \text{for any } 0 \leq s \leq 1. \qquad (2.1.2)$$

But, again, as in the transitivity case (see the remarks following eq.(1.2.4)), because of the basic undesirable consequences in using probability evaluations of the material conditonal operator as interpretations of probability evaluations of conditional expressions, such an extension is not satisfactory. On the other hand, just as in the transitivity case (again, see Figure 1 of Section 1.2), in general, given only the information P(a|b) = s, for even s quite close to unity, we can find particular probability measures P such that P(b′|a′) is close to, or actually zero. Thus, once more, we are led to seek if a second order probability approach to this quandary produces a more reasonable result. Specifically, we ask what is the P-averaged value of P(b′|a′) for given constraint P(a|b) = s and if that evaluation approaches unity as s approaches unity.

In the case of positive conjunction, the premise involved in natural language form is {"if b, then a", "if c, then a"} and the (potential) conclusion is "if b and c, then a". Once again, as in the

transitivity and contraposition cases, the classical logic interpretation via the material conditional operator in boolean form is easily seen to produce the inequality

$$(b \Rightarrow a)\&(c \Rightarrow a) = b'c' \vee b'ac \vee c'ab \;\leq (bc \Rightarrow a), \tag{2.1.3}$$

whence, analogous to the use of the FHH lower bound to produce eq.(1.2.4),

$$[P(b \Rightarrow a) \geq s,\, P(c \Rightarrow a) \geq t\,] \text{ implies } [P(bc \Rightarrow a) \geq s + t - 1], \quad \text{for any } ½ \leq s, t \leq 1. \tag{2.1.4}$$

But, once more, as in both the transitivity and contraposition cases, the potential value of the result in eq.(2.1.4) is diminished by the difficulties involving the probability of the material conditional. And, once more, it can be shown that probabilities P exist for which $P(a|b) = s$ and $P(a|c) = t$, with s and t reasonably high, yet $P(a|\, bc)$ quite low or even zero. Thus, again we are led to seek if a P-averaged evaluation of $P(a|\, bc)$ subject to these constraints produces a more reasonable result.

Finally, we mention the property of *strengthening*, where in a natural language setting the premise is “if b, then a” and the (potential) conclusion is “if b and c, then a”. The classical logic counterpart, using again the material conditional interpretation of “if-then”, is seen to be valid via the simple inequality in boolean form

$$(b \Rightarrow a) = b' \vee a \;\leq\; b' \vee c' \vee a \;= (bc \Rightarrow a), \tag{2.1.5}$$

which if P were to be applied to both sides of eq.(2.1.4), via the standard monotonicity property of probability, produces the equally simple relation

$$[P((b \Rightarrow a) \geq s] \text{ implies } [P(bc \Rightarrow a) \geq s]\text{, for all } s, \quad 0 \leq s \leq 1. \tag{2.1.6}$$

However, as in the cases of transitivity, contraposition, and positive conjunction, the same pattern of difficulty holds: First, eq.(2.1.6) is of limited valued due to the general difficulty in using the probability of the material conditional operator approach to modeling uncertain conditioned information; second, “counterexamples” can be found for which $P(a|b)$ is quite high, yet $P(a|\, bc)$ is very low or zero. Thus, yet again, we whether by suitably averaging $P(a \mid bc)$ over possible P’s, subject to the constraint $P(a|b) = s$, we can obtain a computable function of s which approaches unity as s does.

Table 1 not only provides a summary of closed-form computations for P-averaged conclusion probabilities for transitivity, contraposition, positive conjunction, and strengthening, but also for a number (about thirty total) of other properties (desired and undesired) of logical systems in a second order (uniform prior) probability framework. In the case of the four examples discussed so far, all of them yield P-averaged conclusions that do indeed approach unity as threshold(s) s and/or t approach unity, compatible with commonsense reasoning.

2.2 *Generalizations*

In order to put the above ideas on a more rigorous and general basis, consider the following: Begin with a probability space $(\Omega, \mathcal{B}, P)$, with now P arbitrarily variable and given events a_j, b_j, c,

d in *B*, for which we assume, without loss of generality (attempting to avoid trivial results), that $\emptyset < a_j < b_j$, all j in J, and $\emptyset < c < d$. We also consider a given *premise set* $(a_j|b_j)_{j \text{ in } J}$, where, formally, each "conditional event" $(a_j|b_j)$ corresponds to the conditional expression "if b_j, then a_j", *before being evaluated* via any choice of (well defined) P as $P(a_j|b_j)$, etc., or in multivariable notation, $(a|b)_J$ forms the premise set, while, for purpose of simplicity, the single conditional event (c|d) forms the *potential conclusion*. In turn, we seek to determine two types of *mean conclusion functions*: $\text{meanconc}_i((a|b)_J, (c|d)): A \to [0,1]$, i = 1, 2, where A is a subset of $[0,1]^J$, such as $(\text{open interval}(1/2,1))^J$, and for each choice of $t_J = (t_j)_{j \text{ in } J}$ in A,

$$\text{meanconc}_1((a|b)_J; (c|d))(t_J) =^d E_P(P(c|d) \mid P(a|b)_J = t_J) , \qquad (2.2.1)$$

$$\text{meanconc}_2((a|b)_J; (c|d))(t_J) =^d E_P(P(c|d) \mid P(a|b)_J \geq t_J) , \qquad (2.2.2)$$

where symbolically $P(a|b)_J = t_J$ means $P(a_j|b_j) = t_j$, for all j in J, and similarly, $P(a|b)_J \geq t_J$ means $P(a_j|b_j) \geq t_j$, for all j in J. The mean conclusion functions will be fully rigorously determined once we rigorize what was meant earlier by any P, as being "equally likely", when P varies. A natural way to capture this is as follows: Suppose, first that

$$A =^d \{\alpha_1,\ldots,\alpha_{m+1}\} \subseteq B \qquad (2.2.3)$$

is a set of *atoms* of *B with respect to* $((a|b)_J;(c|d))$. That is, the α_j's form a nonvacuous, exhaustive partitioning of Ω, with each α_j in *B*, so that any nonvacuous boolean function $f(V_o((a|b)_J;(c|d))$ of event variables $V_o((a|b)_J;(c|d))$ from the antecedents and consequents of premise set of conditionals $(a|b)_J$ and potential conclusion conditional (c|d), where

$$((a|b)_J;(c|d)) =^d \{a_j, a_j' b_j, b_j' : j \text{ in } J\} \cup \{c, c'd, d'\}, \qquad (2.2.4)$$

can be (uniquely, necessarily) expressed as a disjoint disjunction of α_j's, indicated through use of the index set $I(.) \subseteq \{1,\ldots,m+1\}$ in

$$f(V_o((a|b)_J;(c|d)) = \bigvee_{j \text{ in } I(f(V((a|b)_{J_o};(c|d))} (á_j) \quad . \qquad (2.2.5)$$

To do this it is clearly sufficient to determine whether each possible conjunctive combination of affirmations and negations of all of $V_o((a|b)_J;(c|d))$ is either equal to $\emptyset$ or a disjoint disjunction of α_j's. Note that the coarsest set of atoms/smallest possible set of atoms/set of atoms generated by A with respect to $((a|b)_J;(c|d))$ is precisely that set of all such nonvacuous possible combinations as above and is denoted as

$$A(V_o)((a|b)_J;(c|d)). \qquad (2.2.6)$$

Thus, any choice of P with respect to its evaluations of all possible boolean functions over $V_o((a|b)_J;(c|d))$ is uniquely determined by the evaluation of P at each atom α_j in A. In fact, the vector $\underline{P} = (P(a_j))_{j \text{ in } J}$, for the purpose of further analysis, may now be identified with P and the set of possible P's relevant to any such investigation becomes simply the simplex $\{\underline{P}: 0_J \leq \underline{P} \leq 1_J$ and $\text{sum}(\underline{P}) = 1$, where $\underline{P}$ is otherwise arbitrary$\}$. For convenience, the above simplex being actually of dimension one less than card(J), and therefore singular with respect to J, is replaced

by designating one atom, say, α_{m+1}, and omitting its evaluation from $\underline{P}$ in the above simplex, but still keeping track of it. That is, we consider the m-simplex, replacing $\underline{P}$ by variable X (with m components, the ith component being $x_i = P(\alpha_i)$) and the missing component,

$$x_{m+1} = P(\alpha_{m+1}) = 1 - \mathrm{sum}(X) \tag{2.2.7}$$

and

$$S_m =^d \{X: 0_m \leq \underline{X} \leq 1_m, \mathrm{sum}(X) \leq 1\}, \tag{2.2.8}$$

Then, the concept of P being equally likely becomes equivalent to assuming X can be identified as a random vector uniformly distributed over S_m (which, in turn, determines the behavior of the missing component). In terms of bayesian analysis, this corresponds to choosing a prior probability distribution, which is uniform over S_m. This distribution is a second order one in the sense described previously: it is, in effect, a distribution of probabilities themselves. as implicitly stated above, the set of atoms, *A*, could be chosen to be $\mathtt{A}(V_o)((a|b)_J;(c|d))$. Choice of the most appropriate set of atoms is somewhat arbitrary, but since all results depend on this choice, the simplest and most natural, at times, may be $\mathtt{A}(V_o)((a|b)_J;(c|d))$.

Also, at times, it will be convenient to consider a class of possible priors to choose from, rather than be restricted to just the uniform distribution. One family of distributions over S_m that includes the uniform one, has a natural characterization compatible with the modeling here, and possesses many desirable properties -- including closure with respect to all index disjoint sums and marginals, among others – is the *Dirichlet* family, indicated symbolically as $\mathrm{dir}(\underline{\lambda})$, with (m+1 by 1) parameter vector $\underline{\lambda} > 0_{m+1}$. The parameter vector $\underline{\lambda}$ is associated with the expectation of $\mathrm{dir}(\underline{\lambda})$, (in fact, it is such up to normalization via $\mathrm{sum}(\underline{\tau})$) and prior knowledge, if available, of the expectation can be transformed into a choice of $\underline{\lambda}$. For the special case of the uniform distribution over S_m, $\underline{\lambda} = 1_m$ and all the component (x_i) expectations are identical to 1/(m+1). For more details on properties and characterization of the Dirichlet family, see [Goodman & Nguyen, 1999a; Section 7.7 of Wilks, 1963; Chapter 40 of Johnson & Kotz, 1972].

Returning to the interpretation of type 1 meanconc functions in eq.(2.2.1), in light of the atomic representation of any event and that of any probability measure here, the conditional probabilities involved can be reinterpreted as simple bilinear functions of variable X, provided that the designated atom $\alpha_{m+1} \leq \&b'_J\&d'$: Letting a_j correspond to the m by 1 vector of one-dimensional components being either 1 or 0, depending on whether for the ith component atom $\alpha_i \leq a_j$ or not (i.e., in the latter case, necessarily disjoint from α_j), then

$$P(a_j|b_j) = a_j^T \cdot X / (a_j^T \cdot X + (a_j' b_j)^T \cdot X) \ , \ j \text{ in } J; \quad P(c|d) = c^T \cdot X / (c^T \cdot X + (c'd)^T \cdot X). \tag{2.2.9}$$

Using eq.(2.2.9), each probability constraint ($P(a_j|b_j) = t_j$) in the antecedent of the conditional expectation in eq.(2.2.1) becomes

$$a_j^T \cdot X = t_j \cdot (a_j^T \cdot X + (a_j' b_j)^T \cdot X), \ j \text{ in } J$$

i.e., the jth plane in variable X determined by

$$((1-t_j) \cdot a_j - t_j \cdot (a_j' b_j))^T \cdot X = 0, \ j \text{ in } J. \tag{2.2.10}$$

The counterpart of eq.(2.2.10) for the type 2 meanconc function is

$$((1-t_j)\cdot a_j \ -t_j\cdot(a_j{'}b_j))^T\cdot X \ \geq 0, \ \ j \text{ in } J, \tag{2.2.11}$$

corresponding to the ordinary probability relations

$$P(a_j{'}b_j) \leq ((1-t_j)/t_j)\cdot P(a_j)\ , \ j \text{ in } J. \tag{2.2.12}$$

With all of the above stated, eqs.(2.2.1) and (2.2.2) can be rigorously interpreted, whether for the special case of X being uniform over S_m or the more general case where X is distributed as dir($\underline{\lambda}$) over S_m. As seen in the transitivity case earlier, in addition to determining meanconc($(a|b)_J$; $(c|d))(t_J)$ for all t_J in its domain, the limiting case $\underset{(t_J \uparrow 1_J)}{\text{limit}}(\text{meanconc}_i((a|b)_J;(c|d))(t_J))$ is of interest. When

$$\underset{(t_J \uparrow 1_J)}{\text{limit}}\left(\text{meanconc}_i((a|b)_J;(c|d))(t_J)\right) = 1, \tag{2.2.13}$$

say that (c|d) is *deduced from* $(a|b)_J$ in the *expected* (or *averaged) probability* logical sense i, i = 1, 2. (Bamber [2000] prefers the term “near surety” in developing a related logic.) When this holds, we will write this relation as

$$(a|b)_J \leq_{EPL} (c|d). \tag{2.2.14}$$

3. Probability Estimation Procedures and Their Relation to the Expected Probability Logic Approach: A First Glimpse

While the approach taken here uses the mean conclusion function for determining properties of reasoning schemes in probability in a second order probability sense, other related approaches to the same issues exist. Bamber [2000] has pointed out, at least in the limiting sense, the “rational closure” approaches in [Lehmann & Magidor, 1992] and that of the “system Z” in [Pearl, 1990] essentially coincide with averaged surety deduction of type 2. However, the actual nonlimiting case evaluation of the meanconc function has no counterpart in these two approaches. On the other hand, alternative approaches based on *first order* probability considerations, including those of Adams’ (see again [Adams; 1975, 1996]), where the idea of a *minimal conclusion function* is developed and that of naive maximal entropy, such as elaborated upon by [Rödder, 2000], can produce different limiting, as well as non-limiting antecedent threshold, evaluations as compared to those computed via meanconc. Adams’ minimum conclusion function counterpart of the mean conclusion function is the function minconc($(a|b)_J$;(c|d)): A $\rightarrow$ [0,1], where for any t_J in A $\subseteq [0,1]^J$, for convenience, giving here only the type two counterpart,

$$\text{minconc}_2((a|b)_J, (c|d))(t_J) =^d \ \inf\{P(c|d)\text{: all possible P, } P(a|b)_J \geq t_J\}. \tag{3.1}$$

The standard naive *maximum entropy function* counterpart is the function maxent($(a|b)_J$; (c|d)): A $\rightarrow$ [0,1], where for convenience we give the type 1 counterpart

$$\text{maxent}_1((a|b)_J;(c|d))(t_J) =^{d} \quad P^*(c|d)\,, \tag{3.2}$$

where

$$P^* = \arg\big(\inf\{\text{ent}(P)\text{: all possible } P,\ P(a|b)_J = t_J\}\big)\,, \tag{3.3}$$

with entropy of any P given as usual as (recalling $X^T = (x_1,...,x_m)$, $x_{m+1} = 1 - \text{sum}(X)$),

$$\text{ent}(P) = \sum_{j=1}^{m+1} (-x_j \cdot \log(x_j))\,. \tag{3.4}$$

For a basic application of the basic maximum entropy approach as outlined in eqs.(3.2)-(3.4), where, typically, the method of Lagrange multipliers is employed to seek for the maximum with respect to the constraints, see [Van Fraasen, 1981]. For a criticism of this approach which produces in a sense a "nonintuitive" result as opposed to the result using a second order probability approach, see [Grove *et al.*, 1997] and Goodman & Nguyen's follow-up and generalization of the issue [Goodman & Nguyen, 1999].

Thus, corresponding to eq.(2.7), we can state that (c|d) is *deduced from* $(a|b)_J$ in the minconc sense i, if

$$\underset{(t_J \uparrow 1_J)}{\text{limit}} \big(\text{minconc}_i((a|b)_J;(c|d))(t_J)\big) = 1 \tag{3.5}$$

and we can state that (c|d) is *deduced from* $(a|b)_J$ in the *maxent sense* i, if

$$\underset{(t_J \uparrow 1_J)}{\text{limit}} \big(\text{maxent}_i((a|b)_J;(c|d))(t_J)\big) = 1,\ \ i = 1, 2. \tag{3.6}$$

In Adams' original terminology, minconc (type 2) deduction corresponds to his *high probability deduction*. Adams has also introduced other criteria for valid deduction of $(a|b)_J$ to (c|d). The one of relevance here involves the minconc function evaluated at $t_J = 1_J$ (in this case, type 1 = type 2) which Adams calls *certainty probability* deduction. More specifically, the criterion for (c|d) being certain-probability deduced from $(a|b)_J$ is that

$$\text{minconc}\ ((a|b)_J, (c|d))(\ 1_J) = 1. \tag{3.7}$$

In addition, a basic modification must be noted for Adams' concepts. Apropos to comments made earlier concerning extending the definition of conditional probabilities when the denominator is zero, i.e., the antecedent is assigned zero probability: Adams, in effect, extends minconc to operate on conditional events in such situations by formally defining the corresponding "conditional probability" to be unity. More specifically, in the context of this paper in analyzing Adams' work, we shall apply the term "strong" to the certainty probability deduction and the minconc types 1 and 2 deduction as already provided in eqs.(3.1), (3.5), etc., and "weak" when in such definitions, either $P(b_j) = 0$ in the premise set is allowed in the formal form of $P(a_j|b_j) = 1$ or $P(d) = 0$ is allowed in the formal form of $P(c|d) = 1$. For simplicity, we will use here

$$(a|b)_J \leq_{HPL} (c|d)\,, \quad (a|b)_J \leq_{CPL} (c|d), \tag{3.7}$$

to indicate high probability (usually, of the strong type), certainty probability deduction validity, (usually of the weak type) respectively, of (c|d) from $(a|b)_J$, where, when required to make these distinctions, the subscript letters S, W, respectively, will be prefixed to indicate strong or weak types.

For purpose of completeness, we present in Table 1 (a good part of which has already appeared in [Goodman, 1999]) a compilation of computations of the type 1 mean conclusion function and type 2 minimum conclusion function restricted, for a variety of combinations of premise sets and potential conclusions, including many corresponding to well-known reasoning schemes of classical logic. Note that in particular, not only transitivity (no. 13), but also contraposition (14), positive conjunction (15), and strengthening (16) all fail to be valid in the strong HP sense, but, as stated earlier, are valid in both the averaged surety sense and the certainty-probability sense (in at least the weak sense). Partial documentation for the derivations may be found in [Bamber *et al.*, 2000], employing various integration techniques. The general assumption throughout Table 1 is that the set of atoms here is $\mathcal{A}(V_o)((a|b)_J;(c|d))$, the minimal set of atoms generated by the premise and conclusion antecedents and consequents (see eq.(2.2.6)) and the random probability vector X has a uniform prior over S_m.

Name and Number of Deduction Scheme $(a\|b)_J$ Potent. Deducing (c|d)	**Given Levels of Premises: $P(a\|b)_J \geq t_J$ for $Minconc_2$; $P(a\|b)_J = t_J$ for $Meanconc_1$**	**Potent. Con-clus. (c|d)**	**$Minconc_2((a\|b)_J;(c\|d))(t_J)$**	**$Meanconc_1((a\|b)_J;(c\|d))(t_J)$**	**Valid For CPL ?**	**Valid for EPL ?**	**Valid for HPL ?**
1. Disjunction	P(a|b) = s, P(a|c) = t	$(a\|b\vee c)$	≥ max(s+t-1,0)	≥ max(s+t-1,0)	YES	YES	YES
2. Bayes	P(a|b) = s, P(c|ab) = t	(c|b)	≥ st	≥ st	YES	YES	YES
3. Cautious Monotonicity	P(a|b) = s, P(c|b) = t	(a|bc)	≥ max(s+t-1,0)	≥ max(s+t-1,0)	YES	YES	YES
4.PSCEA Order	P(a|b) = t, for $\emptyset < a < b$, $\emptyset < c < d$	(c|d)	≥ t	≥ t	YES	YES	YES
5. Reflexivity	P(a|b) = t	(a|b)	t	t	YES	YES	YES
6. Cut	P(a|b) = s, P(c| ab) = t	(ac|b)	≥ st	≥ st	YES	YES	YES
7. Exceptions ($a'bc'$, $bc \neq \emptyset$)	P(a | bc) =s, P(a′|b) = t	(c|b)	≥ max(s+t-1,0)	≥ max(s+t-1,0)	YES	YES	YES
8. Equivalance	P(a|b) = s, P(b|a) = t	$a\Leftrightarrow b$	≥ st	(s+t)/[2(s+t-st)]	YES	YES	YES
9. Strict Modus Ponens	P(a|b) = s, P(b) = t	ab	st	st	YES	YES	YES
10.General Modus Ponens	$P(a\| b\vee c)$ = s, P(b) = t	ab	≥ st	st + (1-t)/2	YES	YES	YES
11.Condition. Bounds 1	P(a|b) = t	$b\Rightarrow a$	≥ t	(2+t)/3	YES	YES	YES
12.Condition. Bounds 2	P(ab) = t	b	≥ t	$2\,\frac{-\,t\log(t) - t(1-t)}{(1-t)^2}$	YES	YES	YES

13.Transitivity-Syllogism	$P(a \mid b) = s$, $P(b \mid c) = t$	$(a \mid c)$	0	$\geq st + (1-t)/2 - \frac{(1-s)s(2s-1)(1-t^2)}{1+2t}$	YES	YES	NO
14.Contra-position	$P(a \mid b) = t$	$(b' \mid a')$	0	$1/t + \frac{(1-t)\log(1-t)}{t^2}$	YES	YES	NO
15. Positive Conjunction	$P(a \mid b) = t$, $P(a \mid c) = t$	$(a \mid bc)$	0	$(1+t)/3 + [((1+t)(2-t)/(3t)) \cdot \theta(t)]$, $\theta(t) = (t^2/4)[\log((2-t)/t)]/(1-t) - ((1-t)^2/4) \cdot \log((1+t)/(1-t)$	YES	YES	NO
16.Strengthen Antecedent	$P(a \mid b) = t$	$(a \mid bc)$	0	approx. t (complicated, but in closed-form)	YES	YES	NO
17.Penguin Triangle $abc'd \neq \varnothing$	$P(a \mid b) = r$, $P(b \mid c) = s$, $P(d \mid c) = t$, $P(a'b \mid d) = u$	$(a'b \mid c)$	0	?	YES (weakly)	YES	NO
18.Modified Penguin Triangle	$P(a \mid b) = r$, $P(b \mid c) = s$, $P(d \mid c) = t$, $d \leq a'b$	$(a' \mid c)$	$\geq \max(s+t-1,0)$	$\geq \max(s+t-1,0)$	YES (weakly)	YES	YES (weakly)
19.Consequ. 1	$P(a \mid b) = t$	a	0	$(1+t)/3$	NO	NO	NO
20.Consequ. 2	$P(a \mid b) = t$	b	0	1/3	NO	NO	NO
21.Consequ. 3	$P(a) = t$	$(a \mid b)$	0	$1/2)(1 + g(t))$, $g(t) = [(1-t) \cdot \log(1-t)] / t - (t \cdot \log(t)) / (1-t)$	YES	YES	NO
22. Consequ.4	$P(b) = t$	$(a \mid b)$	0	1/2	NO	NO	NO
23 Nixon Diamond	$P(ab \mid c) = s$, $P(d \mid a) = t$, $P(d' \mid b) = t$	$(d \mid c)$	0	1/2	YES (weakly)	NO	NO
24.Reverse Cond. Bnd. 1	$P(b \Rightarrow a) = t$	$(a \mid b)$	0	$t + \frac{2(1-t)\log(1-t)}{t^2} + \frac{(1-t)(2+t)}{t}$	YES	YES	NO
25.Reverse Cond. Bnd. 2	$P(a \mid b) = t$	ab	0	t/3	NO	NO	NO
26. Abduction	$P(a \mid b) = s$, $P(a) = t$	b	0	If $s \geq t$: $t/(2s)$, If $s < t$: $\frac{t^3 s(1-t)^2}{2(t^2 - 2st + s)^2}$	NO	NO	NO
27. Induction	For $b_j c$ all disj. $\vee(b_J c) < c$: $P(a \mid b_j \& c) = t_j$, $j=1,\ldots,n$;	$(a \mid c)$	0	?	NO	NO	NO
28.Augmented Induction	For $b_j c$ all disj. $\vee(b_J c) < c$: $P(a \mid b_j \& c) = t_j$, $j=1,\ldots,n$; $P(\vee(b_J) \mid c) = s$	$(a \mid c)$	$\geq \Pi(t_J) - (1-s)$	$\geq \Pi(t_J) - (1-s)$	YES	YES	YES
29. Constrained Conjunction	$P(a) = s$, $P(b) = t$	ab	$\max(s+t-1, 0)$	$(1/2)(\min(s, t) + \max(s+t-1, 0))$	YES	YES	YES
30. Constrained Disjunction	$P(a) = s$, $P(b) = t$	$a \vee b$	$\max(s,t)$	$(1/2)(\max(s, t) + \min(s+t, 1))$	YES	YES	YES

Table 1. Tabulation of minconc and meanconc functions and listing of validity-nonvalidity of selected potential deduction schemes with respect to CPL, EPL, and HPL. Assumption here is minimally-generated set of atoms from relevant premise and conclusion antecedents and consequents and uniform prior over the m-simplex of resulting possible probability functions.

We also illustrate in Table 2 (somewhat overlapping with Table 1) briefly how all three functions minconc, meanconc, and maxent can be similar or quite divergent for various potential deduction schemes. (Again, see [Bamber *et al.*, 2000] for documentation.)

Name and Deduction Scheme $(a\|b)_J$ Potent. Deducing (c|d)	**Given Levels of Premises: $P(a\|b)_J \geq t_J$, for minconc$_2$; $P(a\|b)_J = t_J$, or $P(a\|b)_J \geq t_J$ for meanconc**	**Potent. Con-clus. (c|d)**	**Minconc($(a\|b)_J$;(c|d))(t_J)**	**Meanconc($(a\|b)_J$;(c|d))(t_J)**	**Maxent($(a\|b)_J$;(c|d))(t_J)**
Transitivity-Syllogism	P(a|b) = s, P(b|c) = t	(a|c)	0	$\geq$ st + (1-t)/2 - $\frac{(1-s)s(2s-1)(1-t^2)}{1+2t}$	(1+ (2s –1)t)/2
Contra-position	P(a|b) = t	(b′|a′)	0	$1/t + \frac{(1-t)\log(1-t)}{t^2}$	$1/(1+((1-s)/s)^s)$
Disjunctive Syllogism	P(a∨b) = s, P(a′) = t	b	≥ max(s+t-1,0)	s – (1/2)(1-t)	s – (1/2)(1-t)
Moving Term	P(a∨ bc) = s, ab′c′ > ∅	ab ∨ c	0	(4/5)s + (1/3)(1-s)	(4/5)s + (1/3)(1-s)
Simple Lower Bound	P(a) ≥ s , (1> s > ½)	a	s	(1+s)/2	s

Table 2. Comparison of minconc, meanconc, and maxent functions for five selected possible deduction schemes under same assumptions as in Table 1.

In particular, while obviously maxent and meanconc either coincide or are close to each other in the first four cases of Table 2, they differ considerably with respect to the bottom type of possible deduction scheme. Note, in fact their coincidence for both the invalid deduction scheme *moving term* (since the limit as s approaches 1 is 4/5 < 1) and the valid one *disjunctive syllogism*.

4. Additional Analysis of HPL and CPL: Use of a Conditional Event Algebra

Certainly, as just seen through a number of examples in Tables 1 and 2, Adams' two basic reasoning systems HPL and CPL are quite distinct from EPL. In fact, HPL and CPL are *monotonic* logics in the sense that if (c|d) is deduced from $(a|b)_J$ in either the HPL or CPL sense, then for any other collection of conditionals (or unconditionals) $(a|b)_K$ (with probability space $(\Omega,\mathcal{B},P)$ given, P arbitrarily variable and all a_j, b_j, c, d in $\mathcal{B}$, etc., as usual), (c|d) will also be deduced in the same sense by $(a|b)_{J\cup K}$. On the other hand, EPL is a *nonmonotonic* logic in

general, as we shall see later. (For further background on nonmonotonic logics, see, e.g. the text of [Schlechta, 1997].) Nevertheless, despite the differences, there are certain key connections between HPL, CPL and EPL that will be pointed out.
Adams [1975, 1996] has ostensibly shown already a number of relations among not only HPL and CPL, but other logics. This, of course, does not include EPL, for which Bamber in [Bamber, 2000] has shown basic connections. However, the thrust here is to refine these results more to account for the difference between the weak and strong versions of these logics. In addition, this section will show, for the first time, how a certain form of "conditional event algebra" (to be explained) can be used to provide a complete setting for both elegant formulations and derivations of all of the key results.

This section provides only the barest information necessary for the use of conditional event algebra in deriving properties of HPL and CPL. Later, this will also be useful in additional study of EPL. For a much more detailed presentation, see [Goodman & Nguyen, 1995], where also a history and various characterizations are presented for the particular conditional event algebra discussed below.

To begin with, notice that while the conditional probability P(a|b) appears to be the natural measure of uncertainty or reliability corresponding to an inference rule "if b, then a" -- taking into account the discussion in Section 1.2, precluding use of the possible natural alternative form P(b⇒a)) -- unlike the latter, no standard object or "conditional event" exists which can play the role of the formal argument (a|b) of P in the evaluation P(a|b). Or so, it seems. In fact, a further apparent barrier to the existence of such possible conditional events has been supposedly provided in [Lewis, 1976]. Roughly speaking, Lewis' result states that given any nontrivial probability space $(\Omega,\mathcal{B},P)$, with P arbitrary, one cannot have for any $\varnothing < a < b$ in $\mathcal{B}$, at the same time some event, say (a|b) also in $\mathcal{B}$ with the property that

$$\text{For all P over } \mathcal{B} \text{ (with } P(b) > 0), \ P((a|b)) = P(a|b). \tag{4.1}$$

But, this interesting – and readily proven -- result *does not* restrict (a|b) from existing in a space $\mathcal{B}_o$, which is, in an algebraic sense, strictly larger than $\mathcal{B}$, via an isomorphic-isometric correspondence between any a, b, c,... in $\mathcal{B}$ and $(a|\Omega)$, $(b|\Omega)$, $(c|\Omega)$ in $\mathcal{B}_o$. That is, the co-existence of $(a|\Omega)$, $(b|\Omega)$, $(c|\Omega)$,... and (a|b), (c|d),... all in $\mathcal{B}_o$ does not violate Lewis' "triviality result", even though the $(a|\Omega)$, $(b|\Omega)$, $(c|\Omega)$,... are strongly identifiable (isomorphically – isometrically) with corresponding a, b, c,... in $\mathcal{B}$! In fact, the construction of such conditional events may be carried out in the same completely routine manner that events in a product probability space are obtained. (For further proof of the avoidance of Lewis' restriction, see [Goodman, Mahler & Nguyen, 1997], Sections 11.5 and 12.2.2.) The basic product probability space here, $(\Omega_o,\mathcal{B}_o,P_o)$, extending any given probability space $(\Omega,\mathcal{B},P)$, is simply the one formed out of a countable infinity of independent factor spaces, each identical to $(\Omega,\mathcal{B},P)$. In turn, given any a, b in $\mathcal{B}$, (a|b) is nothing more than the formal algebraic analogue of any of its nontrivial numerical evaluations

$$P(a|b) = P(ab) / (1-P(b')) = \sum_{j=0}^{+\infty}\left((P(b'))^j \cdot P(ab)\right), \tag{4.2}$$

where arithmetic sum corresponds to disjoint $\vee$, multiplication to cartesian product $\times$, and subtraction to negation $(.)'$. That is, one can easily show that if we define the disjoint disjunction

$$(a|b) =^{d} \bigvee_{j=0}^{+\infty} [(b')^{j}\times(ab|\Omega)], \quad \Omega_o =^{d} \Omega\times \Omega\times \Omega\times \ldots , \tag{4.3}$$

where $(.)^{j}\times(..)$ indicates ordinary j-fold cartesian product of (.) with itself, followed by the cartesian product of this result with (..), then, from now on, where necessary, indicating all basic order relations and boolean operators over B_o corresponding to the usual ones over B by the addition of subscript $_o$,

$$(a|\Omega) =_o a\times \Omega_o , \tag{4.4}$$

and the basic consistency relation holds

$$P_o((a|b)) = P(a|b), \text{ for all P over } B, \text{ for } P(b) > 0. \tag{4.5}$$

Even more importantly, the well defined existence of such conditional events (a|b), (c|d),... in B_o, for any a, b, c, d,... in B allows for natural extensions of conjunction, disjunction, negation, and, in fact any boolean function, relative to ordinary events in B, to conditional expressions, and in turn, then allow for probability evaluations of such expressions. First, the basic recursive form that all conditional events here must satisfy is

$$(a|b) =_o (ab|\Omega) \vee (b'\times(a|b)) \text{ (disjoint) }. \tag{4.6}$$

Introduce the conditional-like operator $[.|..]:B_o \times B \rightarrow B_o$, where for any A in B_o and b in B, observing also a recursive form (with disjoint disjunctions)

$$[A|b] =^{d} \bigvee_{j=0}^{+\infty} ((b')^{j}\times(A\&_o(b|\Omega)) =_o A\&_o(b|\Omega)) \vee (b'\times [A|b]) \text{ in } B_o, \tag{4.7}$$

where it is noted that for any a, b in B and A in B_o,

$$[A|\Omega] =_o A , \quad [(a|\Omega) \mid b] =_o (a|b). \tag{4.8}$$

The corresponding probability evaluations to eqs.(4.6)-(4.8) are simply for any P, with P(b) > 0,

$$\begin{aligned} P(a|b) = P(ab) + P(b')P(a|b) , \quad P_o([A|b]) &= P_o(A|\,(b|\Omega)) = P_o(A\&_o(b|\Omega)) / P(b) \\ &= P_o(A\&_o(b|\Omega)) + P(b')P_o([A|b]), \end{aligned} \tag{4.9}$$

etc. In turn, utilizing the above recursive forms, it follows, for any a, b, c, d in B,

$$(a|b) =_o (ab \mid b),\ (a|b)' =_o (a'|b) =_o (a'b\ |b), \text{ provided that } b \neq \emptyset, \tag{4.10}$$
$$(a|b)\&_o(c|d) =_o [A \mid b\vee d], \quad A =^{d} (abcd\ |\Omega) \vee (abd'\times(c|d)) \vee (cdb'\times(a|b)), \tag{4.11}$$

whence if $P(b \vee d) > 0$,

$$P_o((a|b)\&_o(c|d)) = P_o(A)/P(b \vee d),\ P_o(A) = P(abcd) + P(abd')P(c|d) + P(cdb')P(a|b). \qquad (4.12)$$

Note the special cases of *modus ponens* and, generalizing the isomorphic imbedding relation from B into B_o, given via $a \rightarrow (a|\Omega)$, for any a in B, for common antecedent b in B (replacing Ω), $b \neq \varnothing$,

$$(ac\,|b)\&_o(bd\,|\Omega) =_o (abcd\,|\Omega);\ (a|b)\&_o(c|b) =_o (ac\,|b),\ (a|b)\vee_o(c|b) =_o (a \vee c\,|b). \qquad (4.12')$$

Since (Ω, B, P_o) is a legitimate probability space where all of the standard laws of boolean (and sigma) algebra, as well as of probability, are satisfied, then, e.g., the standard modular expansion of probability holds

$$P_o((a|b)\vee_o(c|d)) = P_o((a|b)) + P_o((c|d)) - P_o((a|b)\&_o(c|d)) = P(a|b) + P(c|d) - P_o(A)/(P(b \vee d), \qquad (4.13)$$

where A is as in eq.(4.11), etc. More generally, the following important relation is to be noted – also readily derived from the above recursive structure of conditional events:

For any finite index set J, using multivariable notation, for any collection of a_j, b_j in B, j in J,

$$\&_o(a|b)_J =^d \underset{j \text{ in } J}{\&_o} (a_j|b_j) =_o [\theta_\&(a,b;J) \mid \vee(b_J)]\ ; \qquad (4.14)$$

$$\vee(b_J) =^d \underset{j \text{ in } J}{\vee} (b_j)\ ,\ \ \theta_\&(a,b;J) =^d \ \vee_o\big(\gamma(a,b;K,J) \times \&_o(a|b)_{J\neg K}\big) \text{ (disjoint disjunction)}; \qquad (4.15)$$

$$\gamma(a,b;K,J) =^d \underset{j \text{ in } K}{\&} (a_jb_j)\& \underset{i \text{ in } J\neg K}{\&} (b_i')\ ;\ \gamma(a,b;J,J) = \underset{j \text{ in } J}{\&} (a_jb_j)\ ;\ \gamma(a,b;\varnothing,J) = \underset{j \text{ in } J}{\&} (b_j') =^d \&(b')_J\ ; \qquad (4.16)$$

for any index sets $\varnothing \subseteq K \subseteq J$. In turn, the probability evaluation of the conjunction in eq.(4.14) is

$$P_o(\&_o(a|b)_J) = P_o(\theta_\&(a,b;J)) / P(\vee(b_J));\ \ P_o(\theta_\&(a,b;J)) = \sum_{(\varnothing\neq K\subseteq J)}\big(P(\gamma(a,b;K,J))\cdot P_o(\&_o(a|b)_{J\neg K})\big), \qquad (4.16')$$

noting, that as in its algebraic counterpart in eq.(4.15), the full evaluation of $P_o(\&_o(a|b)_J)$ proceeds recursively, where the factors $P_o(\&_o(a|b)_{J\neg K}))$ are evaluated similarly with J replaced by $J\neg K$, until no more than single conditional probability terms appear. Note also the tie-in of the γ's above with the related disjoint expansion of the conjunction of the material conditional

$$\&(b\Rightarrow a)_J =^d \underset{j \text{ in } J}{\&} ((b_j\Rightarrow a_j)) = \underset{(\varnothing\neq K\subseteq J)}{\vee} (\gamma(a,b;K,J)) = \&(b\Rightarrow a)_J\&(\vee(b)_J) \ \vee\ \&(b')_J \text{ (disjoint)}, \qquad (4.17)$$

where

$$\&(b\Rightarrow a)_J\&(\vee(b)_J) = \&(b\Rightarrow a)_J\&(\vee(a)_J) = \underset{j \text{ in } J}{\vee}\big(\gamma(a,b;K,J)\big) \text{ (disjoint)}. \qquad (4.18)$$

Next, it follows that for any *proper* (or nontrivial) conditional events (a|b), (c|d), i.e., $\varnothing < a < b$ in B, $\varnothing < c < d$ in B (again, see [Goodman & Nguyen, 1995]),

$$\begin{aligned}(a|b) \leq_o (c|d) &\text{ iff } (a|b) =_o (a|b)\&_o(c|d) \text{ iff } (c|d) =_o (a|b) \vee_o (c|d)\\ &\text{ iff } (a|b)\&_o(c|d)'^{\circ} =_o \varnothing \quad \text{ iff } (a|b)\&_o(c'|d) =_o \varnothing_o\\ &\text{ iff } (a \leq c \text{ and } b\Rightarrow a \leq d\Rightarrow c) \text{ iff } (a \leq c \text{ and } c'd \leq a'b)\end{aligned}$$

iff $(b \Rightarrow a \ \leq \ (b \vee d) \Rightarrow cd)$ iff $(c'd \vee d'b \ \leq \ a'b)$
iff for all P over *B*, with $P(b), P(d) > 0$, $P(a|b) \leq P(c|d)$, (4.19)

$(a|b) =_o (c|d)$ iff $((a|b) \leq_o (c|d)$ and $(c|d) \leq_o (a|b))$ iff $(a = c$ and $a'b = c'd)$
iff $(a = c$ and $b = d)$
iff for all P over *B*, with $P(b), P(d) > 0$, $P(a|b) = P(c|d)$. (4.20)

In a related direction (see also the analogue in eq.(1.2.6)), any conditional event (a|b) in B_o satisfies

$$(ab \,|\Omega) \leq_o (a|b) \leq_o (b \Rightarrow a \mid \Omega) \text{ and for } P(b) > 0,\ P(ab) \leq P(a|b) \leq P(b \Rightarrow a). \quad (4.20')$$

Note also that eq.(4.3) shows

$$(a|\varnothing) =_o \varnothing_o \,;\ \text{if } b > \varnothing,\ \ (b|b) =_o \Omega_o \text{ (essentially)}, \quad (4.21)$$

the only possible *improper* (or trivial) conditional events, and for any P over *B*,

$$P(b) = 0 \ \text{ implies } \ P_o((a|b)) = 0\,,\ P(b) > 0 \text{ implies } \ P_o((b|b)) = 1, \quad (4.22)$$

etc. (the opposite of Adams' interpretation, recalling the discussion in Section 3 and earlier).

From now on, we will use the term *product space conditional event algebra* (PSCEA) to refer to the product probability space (Ω_o,B_o,P_o) extending (Ω,B,P) in the above isomorphic-isometric sense of any a in *B* corresponding to (a|b) in B_o isomorphically and for any P, $P_o((a|\Omega)) = P(a)$, together with the conditional event-forming structure a, b $\rightarrow$ (a|b), etc.

We will also need to imbed an important operator developed independently in [Adams, 1975, 1996] and in [Calabrese, 1987, 1994], often called "quasi-conjunction", since, as important an operator as it will be seen later it is, it is not only non-boolean in structure, but fails to form a full lattice operation with its DeMorgan dual [Goodman & Nguyen, 1995]. On the other hand, this operator appears to produce a sort of conjunction that at times may be the appropriate interpretation of "and" in a conditional setting. (See, again [Goodman & Nguyen, 1995] for further details.) In any case, the appropriate definition in the PSCEA setting for this non-boolean operator is, for given probability space (Ω,B,P) and any events $\varnothing < a_j < b_j$ in *B*, j in J (any finite index set) producing proper conditional events $(a_j|b_j)$ in B_o, j in J, is simply the *direct* (non-associative) one, using the multivariable notation from eqs.(4.14), (4.16),

$$\&_{AC}(a|b)_J =^d \underset{j\text{ in }J}{\&_{AC}} (a_j|b_j) =^d (\&(b \Rightarrow a)_J \mid \vee(b_J))$$

$$= \big((\&(b \Rightarrow a)_J \,\&(\vee(b_J)) \mid \vee(b_J)\big) \quad \text{in } B_o. \quad (4.23)$$

Note that this version of $\&_{AC}$ in PSCEA is always well defined since its domain consists of proper conditional events and thus the identification property in eq.(4.20) can be used to test equality of other proper conditional events with those produced by $\&_{AC}$, under the conditionas

that the latter produces a proper conditional event. Indeed, the only improper conditional event that $\&_{AC}(a|b)_J$ can assume in general is $\emptyset_o$, since

$$(\&_{AC}(a|b)_J)'^{\circ} = (\vee(a'b)_J \mid \vee(b_J)) \;, \quad \vee(a'b)_J =^d (\underset{j \text{ in } J}{\vee}(a_j'b_j)) \;, \tag{4.24}$$

and since $\vee(b_J) > \emptyset$ always is assumed,

$$\&_{AC}(a|b)_J = \Omega_o \quad \text{iff} \quad (\vee(a'b)_J \mid \vee(b_J)) = \emptyset_o \,,$$

which is impossible, since by the proper conditional event assumption, each $a_j'b_j > \emptyset$, hence $\vee(a'b)_J > \emptyset$. For the case of J ={1}, $\&_{AC}(.)_J$ reduces to the usual identity operator, in common with that of $\&_o(.)_J$ and $\vee_o(.)_J$:

$$\&_{AC}(a|b)_J = \&_{AC}\,(a_1|b_1) = (a_1|b_1). \tag{4.25}$$

Clearly, when comparing the forms in eqs.(4.14), (4.15), (4.18), and (4.23), $\&_{AC}$ and $\&_o$ differ in that each corresponding term of $\&_o$ has also a cartesian product factor. Hence, it readily follows that it is always true that

$$\&_o(a|b)_J \;\le\; \&_{AC}(a|b)_J \,. \tag{4.26}$$

Note that the probability evaluation of $\&_{AC}(a|b)_J$, analogous to the expansion in eq.(4.16′), is

$$P_o(\&_{AC}(a|b)_J) = P(A(a,b;J)) \,/\, P(\vee(b_J)); \;\; P_o(A(a,b;J)) = \sum_{(\emptyset \ne K \subseteq J)} \big(P(\,\gamma(a,b;K,J))\big), \tag{4.27}$$

which, in general, can be considerably larger than the counterpart $P_o(\&_o(a|b)_J)$.

In turn, we next state some important connections between ordering with respect to PSCEA conjunction and AC conjunction. From, now on, for simplicity, we omit subscript o from ordering and equality relations and negation in PSCEA, but retain it for conjunction (and disjunction), in order to distinguish it from $\&_{AC}$, which will be widely used.

Theorem 4.1. Ordering relations between $\&_o$ and $\&_{AC}$

For (Ω,B,P) a given probability space with PSCEA extension (Ω_o,B_o,P_o), P arbitrary, and any proper conditional events $(a_j \mid b_j)$ in B_o, j in J (finite), (c|d) in B_o,

(i) $\&_o(a|b)_J \le (c|d)$ iff $\underset{(\emptyset\ne K\subseteq J)}{\text{Or}}(\,\&_{AC}(a|b)_K \le (c|d)\,)$.

(ii) $\text{not}(\&_o(a|b)_J \le (c|d))$ iff $\underset{(\emptyset\ne K\subseteq J)}{\text{And}}(\,\text{not}(\,\&_{AC}(a|b)_K \le (c|d)\,)\,)$

iff $\&_o(a|b)_J\,\&_o(c'|d) \ne \emptyset_o$

iff $\underset{(\emptyset\ne K\subseteq J)}{\text{And}}(\;\&_{AC}(a|b)_K \ne \emptyset_o)$ and $\underset{(\emptyset\ne K\subseteq J)}{\text{And}}(\;\&_{AC}\big((a|b)_K\,,\,(c'|d)\big) \ne \emptyset_o\,)$.

(iii) $\&_o(a|b)_J = \emptyset_o$ iff $\underset{(\emptyset\ne K\subseteq J)}{\text{Or}}(\,\&_{AC}(a|b)_K = \emptyset_o)$.

(iv) $\&_o(a|b)_J \ne \emptyset_o$ iff $\underset{(\emptyset\ne K\subseteq J)}{\text{And}}(\,\&_{AC}(a|b)_K \ne \emptyset_o\,)$.

Proof: (i) is the same as Lemma 10 in [Bamber *et al.*, 2000].

(iii) is straightforward via cases of card(J) = 1, 2, 3. One can procede analogously for higher values of card(J).

(ii) and (iv) follow logically from (i) and (iii). ■

5. Additional Analysis of HPL and CPL: Use of Algebraic Characterizations of Particular Probability Relations

While in Section 4 a number of algebraic properties were developed directly related to PSCEA, in this part, relations are exhibited for the most part between algebraic descriptions and probability bounds. This is motivated by the following definitions, which refine and carefully delineate between the weak and strong versions of HPL and CPL. More specifically, letting, as usual, (Ω,B,P) be a given probability space with probability measure $P:\mathit{B}\to[0,1]$ arbitrarily variable, $(\Omega_o,\mathit{B}_o,P_o)$ its PSCEA extension, J a finite index set and events $\varnothing < a_j < b_j$ in B, j in J, $\varnothing < c < d$ in B, using the multivariable notation as previously developed whenever possible so that in the following, e.g., $(a|b)_J$ represents the premise set of conditional events $(a_j|b_j)$ (or unconditional/ ordinary events whenever $b_j = \Omega$) and $(c|d)$ represents the single (for purpose of simplicity) potential conclusion conditional event (or unconditional event, when $d = \Omega$).

By convention, let us call the collection of above assumptions, *Basic Assumption I*.

For general background, we again refer to [Adams, 1966, 1975, 1986, 1996; Goodman & Nguyen, 1998; Goodman, 1999; and Bamber *et al.*, 2000].

Definition 1. Say that *strong high probability deduction (or logic) (SHPL)* holds with respect to $((a|b)_J;(c|d))$ (or that $(a|b)_J$ *deduces* $(c|d)$ *in the SHPL sense*, etc.), written symbolically as

$$(a|b)_J \leq_{SHPL} (c|d), \tag{5.1}$$

iff

$$\lim_{t_J\uparrow 1_J}\left(\text{minconc}((a|b)_J;(c|d))(t_J)\right) = 1, \tag{5.2}$$

i.e.,

$$\begin{gathered}(\text{for any } 0 < \varepsilon < 1)(\text{there is a } 0 < \delta_\varepsilon < 1)(\text{for any } P)\\ \big(\text{if } [P(a|b)_J \geq 1-\delta_\varepsilon)], \text{ then } [P(c|d) \geq 1-\varepsilon]\big),\end{gathered} \tag{5.3}$$

noting all conditional probabilities are in the ordinary sense, i.e., $P(b)_J > 0_J$.

Definition 2. Say that *strong high probability (SHPL) consistency* holds with respect to $(a|b)_J$ iff the "if-part" of eq.(5.3) is nonvacuously satisfied for all possible threshold levels, i.e.,

$$(\text{for any } 0 < \delta < 1)(\text{ there is a } P_\delta)(\ P_\delta(a|b)_J \geq 1-\delta), \tag{5.4}$$

Definition 3. Say that *weak high probability deduction (or logic) (WHPL)* holds with respect to $((a|b)_J;(c|d))$, (or that $(a|b)_J$ *deduces* $(c|d)$ *in the WHPL sense*, etc.), written symbolically as

$$(a|b)_J \leq_{WHPL} (c|d) \tag{5.5}$$

iff in the expressions in eqs.(5.2) or (5.3) we allow possibly some of the $P(b_j)$ to be 0 and *formally* interpret $P(a_j|b_j) = 1$, and similarly for $P(c|d)$, i.e.,

$$\begin{aligned}&(\text{for any } 0 < \varepsilon < 1)(\text{there is a } 0 < \delta_\varepsilon < 1)(\text{for any } P)\\ &\big(\text{if } [(\text{for each } j \text{ in } J)(\text{either } P(a_j|b_j) \geq 1\text{-}\delta_\varepsilon \text{ or } P(b_j) = 0)], \text{ then}\\ &\quad [\text{either } P(c|d) \geq 1\text{-}\varepsilon \text{ or } P(d) = 0]\big). \qquad (5.6)\end{aligned}$$

Definition 4. Say that *weak high probability (WHPL) consistency* holds with respect to $(a|b)_J$ iff the "if-part" of eq.(5.6) is nonvacuously satisfied for all possible threshold levels, i.e.,

$$(\text{for any } 0 < \delta < 1)(\text{ there is a } P_\delta)(\text{ for each } j \text{ in } J)(\text{either } P_\delta(a_j|b_j) \geq 1\text{-}\delta \text{ or } P_\delta(b_j) = 0). \qquad (5.7)$$

Definition 5. Say that *strong certainty probability (SCPL) deduction (or logic)* holds with respect to $((a|b)_J;(c|d))$ (or that $(a|b)_J$ *deduces* $(c|d)$ *in the SCPL sense*, etc.), written symbolically as

$$(a|b)_J \leq_{SCPL} (c|d), \qquad (5.8)$$

iff

$$\mathrm{minconc}((a|b)_J;(c|d))(1_J)) = 1, \qquad (5.9)$$

i.e.,

$$(\text{for any } P)\big(\text{if } [P(a|b)_J =1], \text{ then } [P(c|d) =1]\big). \qquad (5.10)$$

Definition 6. Say that *strong certainty probability (SCPL) consistency* holds with respect to $(a|b)_J$ iff the "if-part" of eq.(5.10) is nonvacuously satisfied at threshold level 1, i.e.,

$$(\text{there exist } P)(\ P(a|b)_J =1). \qquad (5.11)$$

Definition 7. Say that *weak certainty probability (WCPL) deduction (or logic)* holds with respect to $((a|b)_J;(c|d))$ (or that $(a|b)_J$ *deduces* $(c|d)$ *in the WCPL sense*, etc.), written symbolically as

$$(a|b)_J \leq_{WCPL} (c|d), \qquad (5.12)$$

iff in the expressions in eqs.(5.9) or (5.10) we allow possibly some of the $P(b_j)$ to be 0 and *formally* interpret $P(a_j|b_j) = 1$, and similarly for $P(c|d)$, i.e.,

$$\begin{aligned}&(\text{for any } P)\big(\text{if } [(\text{for each } j \text{ in } J)(\text{either } P(a_j|b_j) = 1 \text{ or } P(b_j) = 0)], \text{ then}\\ &\quad [\text{either } P(c|d) = 1 \text{ or } P(d) = 0]\big). \qquad (5.13)\end{aligned}$$

Definition 8. Say that *weak certainty probability (WCPL) consistency* holds with respect to $(a|b)_J$ iff the "if-part" of eq.(5.13) is nonvacuously satisfied at threshold level 1, i.e.,

$$(\text{there exists } P)(\text{ for each } j \text{ in } J)(\text{either } P(a_j|b_j) = 1 \text{ or } P(b_j) = 0). \qquad (5.14)$$

Remarks. By the very definitions above, it follows immediately that:

(i) SHPL consistency implies WHPL consistency.

(ii) SCPL consistency implies WCPL consistency.
(iii) The remaining relations among the various concepts will be examined below.
The next results allow us to determine for the most part algebraically when any of the above concepts hold true.

Theorem 5.1 Characterization of SHPL consistency. (Extension of [Adams, 1975])

Under Basic Assumption I, the following statements are equivalent:

(i) SHPL consistency holds with respect to $(a|b)_J$.

(ii) $\&_o(a|b)_J \neq \emptyset_o$.

(iii) $\underset{(\emptyset\neq K\subseteq J)}{\text{And}}\,(\&_{AC}(a|b)_K \neq \emptyset_o)$.

(iv) There is a positive integer M and an exhaustive nonvacuous partitioning $\{K_1,...,K_M\}$ of J such that, using the notation of eq.(4.16), letting

$$K(j) =^d \bigcup_{i=1}^{j-1}(K_i)\ ,\ j=1,2,...,M; \qquad K(0) =^d \emptyset\ , \tag{5.15}$$

$$\gamma(a,b;K_j,J\neg K(j)) \neq \emptyset,\ \text{ for } j = 1,..., M\ . \tag{5.16}$$

Note that necessarily for the Mth term, $K(M) = \emptyset$ and

$$\gamma(a,b;K_M,\emptyset) = \&(a_{K_M}) \neq \emptyset. \tag{5.17}$$

(v) There is a positive integer M and an exhaustive nonvacuous partitioning $\{K_1,...,K_M\}$ of J such that, using the notation of eq.(5.15),

$$\&_{AC}(a|b)_{J\neg K(j)} \neq \emptyset, \quad \text{for } j = 0, 1,..., M\text{-}1.$$

(vi) $\underset{(\emptyset\neq K\subseteq J)}{\text{And}}\,(\underset{\emptyset\neq L\subseteq K)}{\text{Or}}\,(\gamma(a,b;L,K) \neq \emptyset))$.

(vii) $\underset{(\emptyset\neq K\subseteq J)}{\text{And}}\big(\&(b \Rightarrow a)_K\ \&\ (\vee(a)_K) \neq \emptyset\big)$.

Proof: (iii) iff (vi) iff (vii): This follows immediately from definition of $\&_{AC}$, etc.

(iv) iff (v) holds due to the basic structure of $\&_{AC}$ (see Section 4).

(ii) iff (iii) follows directly from Theorem 4.1 (iv).

(ii) implies (iv): Using eqs.(4.14)-(4.16), (ii) shows that for some $\emptyset \neq K_1 \subseteq J$, there is a term $\gamma(a,b;K_1,J) \times \&_o(a|b)_{J\neg K_1} \neq \emptyset_o$ in the consequent $\theta_\&(a,b;J)$ of $\&_o(a|b)_J$. Thus, both sides of the cartesian product must be non-null, whence $\&_o(a|b)_{J\neg K_1} \neq \emptyset$. But, next, replacing J in the above reasoning process by $J\neg K_1$, we next obtain some $\emptyset \neq K_2 \subseteq J\neg K_1$ and a term $\gamma(a,b;K_2,$

$J\neg K_1\neg K_2) \times \&_o(a|b)_{J\neg K_1\neg K_1} \neq \emptyset_o$. We continue the process until some M is found so that (guaranteed) $J\neg K_1 \ldots \neg K_M = \emptyset$.

(iv) implies (i): This follows the guidelines of Adams' approach [1996], where since it is readily verified that the $\gamma(a,b;K_j,J\neg K(j)) \neq \emptyset$, for j = 1,..., M , are all mutually disjoint subsets of $\vee(b)_J$, for any given real d, $0 < \delta < 1$, we construct a P_δ by assigning

$$P_\delta(\gamma(a,b;K_j,J\neg K(j)) =^d \delta^{j-1} - \delta^j \text{ , for } j = 1,\ldots, M\text{-}1 \text{ ;}$$
$$P_\delta(\gamma(a,b;K_M,\emptyset) \quad =^d \delta^{M-1} \quad \text{, for } j = M. \tag{5.18}$$

Then, since for any K_j, and all i in K_j, b_i is disjoint from $\&(b')_{J\neg K(j-1)}$, the latter being $\geq$

$\bigvee_{k=1}^{j-1}(\gamma(a,b;K_k,J\neg K(k)))$, it follows that

$$P_\delta(b_i) \leq 1 - P_\delta\Big(\bigvee_{k=1}^{j-1}(\gamma(a,b;K_k,J\neg K(k)))\Big) = 1 - \sum_{k=1}^{j-1}(\delta^{k-1} - \delta^k) = \delta^{j-1},\ j = 1,\ldots,M. \tag{5.19}$$

On the other hand, since for any i in K_j,

$$a_i \geq \&(a)_{K_j} \geq \gamma(a,b;K_j,J\neg K(j)) > \emptyset,$$

$$P_\delta(a_i) \geq P_\delta(\gamma(a,b;K_j,J\neg K(j)), \text{ for } j = 1,\ldots, M. \tag{5.20}$$

Thus, combining eqs.(5.18)-(5.20) shows, for all i in K_j,

$$P(a_i|b_i) \geq (\delta^{j-1} - \delta^j) / \delta^{j-1} = 1\text{-}\delta, \text{ for } j = 1,\ldots, M. \tag{5.21}$$

(i) implies (ii): Since all laws of probability apply to PSCEA, the FHH lower bound in eq.(1.1.5) holds applied to unconditionals replaced by conditionals and P by P_o, i.e., assuming $P(b_j) > 0$, all j in J,

$$\max\Big(\sum_{j \text{ in } J}(P(P(a_j|b_j) - (\mathrm{card}(J)-1), 0\Big) \leq P_o(\&_o(a|b)_J) \leq \min_{j \text{ in } J}(P(a_j|b_j)). \tag{5.22}$$

(i) then implies for every real 0 < d < 1, there is a P_δ with $P_\delta(a_j|b_j) \geq 1\text{-}\delta$, for all j in J, which combined with eq.(5.22), for all δ, $0 < \delta < 1/\mathrm{card}(J)$, shows

$$0 < 1 - \delta\cdot\mathrm{card}(J) = \mathrm{card}(J)\cdot(1\text{-}\delta) - (\mathrm{card}(J) - 1) \leq \sum_{j \text{ in } J}(P(a_j|b_j)) - (\mathrm{card}(J) - 1)$$
$$\leq P_o(\&_o(a|b)_J), \tag{5.23}$$

which certainly implies that $\&_o(a|b)_J \neq \emptyset_o$. ■

Theorem 5.2. Characterization of SHPL deduction. (Extension of [Adams, 1975])

Under Basic Assumption I and SHPL consistency, the following statements are equivalent:

(i) $(a|b)_J \leq_{SHPL} (c|d)$.

(ii) $\&_o(a|b)_J \leq (c|d)$.

(iii) $\underset{(\emptyset \neq K \subseteq J)}{\mathrm{Or}} (\&_{AC}(a|b)_K \leq (c|d))$.

Proof: By Theorem 4.1(i), (ii) iff (iii).

not(ii) implies not(i): Suppose not(ii). By Theorem 4.1 (ii), not(ii) is equivalent to

$$\&_o(a|b)_J \&_o(c'|d) \neq \emptyset_o. \tag{5.24}$$

But, by Theorem 5.1, where we assume o not in J, letting $(a_o|b_o) =^d (c'|d)$ and $J_o =^d J \cup \{o\}$, replacing J there by J_o:
For each real δ, $0 < \delta < 1$, there is a P_δ such that for all j in J_o, $P(a_j|b_j) \geq 1-\delta$,
i.e., $P_\delta(a_j|b_j) \geq 1-\delta$, for all j in J, and $P_\delta(c'|d) \geq 1-\delta$, i.e., $P(c|d) \leq \delta$. (5.25)

Thus, the results in eq.(5.25) clearly shows not(i).

(ii) implies (i): Simply apply the FHH inequality in a PSCEA setting, as in eqs.(5.22), (5.23). ■

Theorem 5.3. Characterization of WHPL consistency.

Under Basic Assumption I, the following statements are equivalent:

(i) WHPL consistency holds with respect to $(a|b)_J$.

(ii) $\&(b \Rightarrow a)_J \neq \emptyset$.

(iii) $\vee(a'b)_J \neq \Omega$.

Proof: (ii) iff (iii) is immediate.

(ii) implies (i): From eq.(4.17), (ii) implies there is some K, $\emptyset \neq K \subseteq J$, so that $\gamma(a,b;K,J) \neq \emptyset$. Then, pick any P such that $P(\gamma(a,b;K,J)) = 1$. This immediately implies that (i) is satisfied.

not(ii) implies not(i): Suppose not(ii). First, consider any probability measure P and $K_P =^d \{j \text{ in } J: P(b_j) = 0\}$. If $K_P = J$, then $P(\vee(b)_J) \leq \Sigma(P(b_J)) = 0$, implying $P(\vee(b)_J) = 0$. But, by not(ii), $\&(b \Rightarrow a)_J = \emptyset_o$ and hence $\vee(a'b)_J = \Omega$, implying $\vee(b)_J = \Omega$, contradicting the above probability evaluation. Thus, we must always have $\emptyset \subseteq K_P \subset J$ (proper), and hence, $\emptyset \neq J \neg K_P$. In turn, note that by the definition of K_P,

$$1 = P(\Omega) = P(\vee(a'b)_J) = P(\vee(a'b)_{J \neg K_P}). \tag{5.26}$$

Now, suppose (i) were true. In particular, choose any real δ, $0 < \delta < 1/(1+\mathrm{card}(J))$ and any P_δ satisfying $P_\delta(a_j|b_j) \geq 1-\delta$ or $P(b_j) = 0$, for any j in J. Thus, $P_\delta(b_j) = 0$, for all j in K_P and $P(a_j|b_j) \geq 1-\delta$, for all j in $J\neg K_P$. The latter is the same as

$$P_\delta(a_j'b_j) \leq (\delta/(1-\delta))P(a_j), \text{ for all j in } J\neg K_P. \tag{5.27}$$

Combining eqs.(5.26) and (5.27),

$$1 = P(\vee(a'b)_{J\neg K_P}) \leq \Sigma(P_\delta(a'b)_{J\neg K_P}) \leq (\delta/(1-\delta))\cdot\Sigma(P_\delta(a)_{J\neg K_P})$$
$$\leq (\delta/(1-\delta))\cdot\mathrm{card}(J\neg K_P) \leq (\delta/(1-\delta))\cdot\mathrm{card}(J) < 1,$$

a contradiction. Hence, not(i) must hold. ■

Theorem 5.4. Under SHPL consistency, WHPL deduction implies SHPL deduction.

Under Basic Assumption I and SHPL consistency,

$$(a|b)_J \leq_{WHPL} (c|d) \quad \text{implies} \quad (a|b)_J \leq_{SHPL} (c|d).$$

Proof: Suppose not (SHPL). Then, eq.(5.25) in the proof of Theorem 5.2 clearly shows a violation of both SHPL and WHPL. ■

Theorem 5.5.

Under the Basic Assumption I, if $(a|b)_J$ is SHPL consistent and $(a|b)_J \leq_{SHPL} (c|d)$, then

$$\&((b\Rightarrow a)_J) \leq d\Rightarrow c. \tag{5.28}$$

Proof: By hypothesis, using Theorem 5.1,

$$\varnothing \neq \&_o(a|b)_J \leq (c|d). \tag{5.29}$$

Hence, by Theorem 4.1(iv), (i),

$$\underset{(\varnothing\neq K\subseteq J)}{\mathrm{And}}(\&_{AC}(a|b)_K \neq \varnothing) \text{ and (there exists } K_1)(\varnothing\neq K_1\subseteq J)(\&_{AC}(a|b)_{K_1} \leq (c|d)). \tag{5.30}$$

Now, as pointed out in Section 4, under Assumption I, for any K, $\varnothing\neq K\subseteq J$, $\&_{AC}(a|b)_K \neq \Omega_o$. On the other hand, the left-hand side of eq.(5.30) shows neither can any $\&_{AC}$ result be null either, i.e., we must have, in particular, $\&_{AC}(a|b)_{K_1}$ being a proper conditional event. Hence, the basic ordering criterion for PSCEA in eq.(4.19) can be invoked to characterize the right-hand side of eq.(5.30). Thus, noting again from Section 4, the structure of $\&_{AC}$, we obtain from the right-hand side of eq.(5.30)

$$\&_{AC}(a|b)_{K_1} =^{w} (A|B) \leq (c|d) \text{ iff } [A \leq c \text{ and } B\Rightarrow A \leq d\Rightarrow c], \tag{5.31}$$

where

$$A =^{d} \&((b\Rightarrow a)_{K_1})\& B\ ,\ \ B =^{d} \vee(b_{K_1}).$$

In turn, part of the right-hand side of eq.(5.31) implies

$$\&(b\Rightarrow a)_J \ \leq\ \ \&(b\Rightarrow a)_{K_1} = B\Rightarrow A\ \leq\ d\Rightarrow c,$$

the desired result. ■

Theorem 5.6. Characterization of WHPL deduction. (a variation of [Adams, 1986, 1996])

Under Basic Assumption I and WHPL consistency, the following statements are equivalent:

(i) $(a|b)_J \leq_{WHPL} (c|d)$.

(ii) $\underset{\varnothing\neq K\subseteq J)}{\mathrm{Or}}\big([\&(b\Rightarrow a)_K \ \&\ (\vee(b)_k \leq c]\ \text{and}\ \ [\&(b\Rightarrow a)_K \leq (d\Rightarrow c)]\big)$.

(iii) $\underset{\varnothing\neq K\subseteq J)}{\mathrm{Or}}\Big(\ [(\&(b\Rightarrow a)_K \ \&\ (\vee(b)_k = \varnothing)\ \text{ and } (\&(b\Rightarrow a)_K \leq (d\Rightarrow c))]$

$\qquad\qquad \text{or}\ \ \ [\varnothing \neq \&_{AC}(a|b)_K \leq (c|d)]\ \Big)$

Proof: [Adams, 1986] provides an equivalent, but different-appearing formulation and proof of the above theorem. A complete self-contained proof here is given in Appendix A. ■

Remark 1.

Improving upon Theorem 5.4, Theorem 5.6 shows directly that under $(a|b)_J$ having SHPL consistency, WHPL and SHPL deduction of $(c|d)$ from $(a|b)_J$ are equivalent. It also shows directly that under WHPL consistency, WHPL deduction implies WCPL deduction of $(c|d)$ from $(a|b)_J$. See also the summary of consistency and deduction relations in Theorem 5.11.

Remark 2.

As an illustration of weak vs. strong HPL deduction, consider, e.g., the case of $J = \{1,2\}$, where for purpose of nontriviality, we assume that $(a|b)_J$ is not SHPL consistent, but is WHPL consistent, i.e., from the above theorems,

$$\&_o(a|b)_J = \varnothing_o \ \ \text{ and }\ \ \&(b\Rightarrow a)_J \neq \varnothing. \tag{5.32}$$

Using eqs.(4.11), (4.16), (4.17), it follows that eq.(5.32) (under our basic Assumption I) is equivalent to

$$a_1a_2 = a_1b_2' = a_2b_1' = \varnothing,\quad b_1'b_2' \neq \varnothing. \tag{5.33}$$

Consider next the possible situations with respect to any real δ, $0 < \delta < 1$ and P_δ such that the weak consistency condition at δ is satisfied, i.e., either $P_\delta(a_j|b_j) \geq 1-\delta$ or $P_\delta(b_j) = 0$, $j = 1, 2$.

Case 1: $P_\delta(a_j|b_j) \geq 1-\delta$, j = 1, 2. But, by appealing to the FHH lower bound (eq.(5.22)), it is clear that for d sufficiently small, we would have $(P_\delta)_o((a_1|b_1)\&_o(a_2|b_2))$ large, contradicting the left-hand side of eq.(5.32), where the value should be zero.
Case 2: $P_\delta(a_1|b_1) \geq 1-\delta$ and $P_\delta(b_2) = 0$. But, eq.(5.33) shows that since $P_\delta(b_2') = 1$, $P_\delta(a_1) = P_\delta(a_1b_2') = P_\delta(\emptyset) = 0$, contradicting $P_\delta(a_1|b_1) \geq 1-\delta$ above.
Case 3: $P_\delta(a_2|b_2) \geq 1-\delta$ and $P_\delta(b_1) = 0$. This yields, dually, the same contradiction as in Case 2.
Case 4: This is the only remaining possibility: $P_\delta(b_1) = P_\delta(b_2) = 0$, i.e., $P_\delta(b_1'b_2') = 1$.

Thus, so far,

$$(a|b)_{WHPL} \leq (c|d) \text{ iff (for all real } \varepsilon)(0<\varepsilon<1)(\text{there is a real } \delta)(0<\delta<1)(\text{for all P}) \\ (\text{if } P(b_1'b_2') = 1, \text{ then either } P(c|d) \geq 1-\varepsilon \text{ or } P(d) = 0). \quad (5.34)$$

Situation 1: $b_1'b_2'c'd \neq \emptyset$. But, pick any P such that $P(b_1'b_2'c'd) = 1$, thus showing not($(a|b)_J \leq_{WHPL} (c|d)$) here.

Situation 2: $b_1'b_2' \leq d \Rightarrow c$, the only remaining possibility, which obviously from the constraint on possible P's works, i.e., for any P satisfying the "if-part" of eq.(5.34), $P(d\Rightarrow c) = 1$, i.e., either $P(d) = 0$ or $P(c|d) = 1$.

Hence, in summary, for J = {1,2} and $(a|b)_J$ being WHPL consistent, but not SHPL consistent,

$$(a|b)_J \leq_{WHPL} (c|d) \text{ iff } b_1'b_2' \leq d\Rightarrow c\,, \quad (5.35)$$

with no proper conditional probabilities involved when the deduction holds.

Lemma 5.1 A useful characterization.

Here, assume a probability space (Ω,B,P) present, with P variable, J a finite index set, $\emptyset \neq e_j$, f in *B*, j in J. Then, the following two statements are equivalent:

(i) (There is a P)(P(f) = 1 and for all j in J, $P(e_j) > 0$).

(ii) $\underset{j \text{ in } J}{\text{And}}(\ e_j f \neq \emptyset\ )$.

Proof: Straightforward. ■

Theorem 5.7. Characterization of WCPL consistency.

Under the Basic Assumption I, the following statements are equivalent:

(i) $(a|b)_J$ is WCPL consistent.

(ii) $\&(b\Rightarrow a)_J \neq \emptyset$.

Proof: (ii) implies (i): Suppose (ii) holds. Let P be such that $P(\&(b\Rightarrow a)_J) = 1$. This yields $P(b_j\Rightarrow a_j) = 1$, and hence $P(a_j|b_j) = 1$ or $P(b_j) = 0$, as required, for all j in J.
not(ii) implies not(i): Suppose not(ii) holds. Then, it is impossible to find any P such that for all j in J, either $P(b_j) = 0$ or $P(a_j|b_j) = 1$, since this would imply $P(b_j\Rightarrow a_j) = 1$ and hence $P(\&(b\Rightarrow a)_J) = 1$, implying $(\&(b\Rightarrow a)_J) \neq \emptyset$, contrary to the assumption. ■

Theorem 5.8. Characterization of WCPL deduction. [Adams, 1996]

Under the Basic Assumption I and WCPL consistency for $(a|b)_J$, the following statements are equivalent:

(i) $(a|b)_J \leq_{WCPL} (c|d)$.

(ii) $(\&(b\Rightarrow a)_J) \leq d\Rightarrow c$.

Proof: The proof is basic and has appeared in a number of places. However, for completeness, we present a brief outline.

(ii) implies (i): Suppose (ii). Then for any P such that $P(b_j) = 0$ or $P(a_j|b_j) = 1$, i.e., $P(b_j\Rightarrow a_j) = 1$, all j in J, implying $1 = P((\&(b\Rightarrow a)_J) \leq P(d\Rightarrow c)$.

not(ii) implies not(i): Suppose not(ii). Then, $(\&(b\Rightarrow a)_J)\&c'd \neq \emptyset$, and by choosing P, such that $P((\&(b\Rightarrow a)_J)\&c'd) = 1$, we easily see that not(i) holds. ■

Theorem 5.9. Characterization of SCPL consistency.

Under Basic assumption I, the following statements are equivalent:

(i) $(a|b)_J$ is SCPL consistent.

(ii) $\underset{i \text{ in } J}{\text{And}}(\ b_i\& (\&(b\Rightarrow a)_J) \neq \emptyset)$.

(iii) $\underset{i \text{ in } J}{\text{And}}(\ a_i\& (\&(b\Rightarrow a)_J) \neq \emptyset)$.

(iv) $\underset{i \text{ in } J}{\text{And}}(\ a_i\neg (\&(b\Rightarrow a)_J) \neq \emptyset)$.

Proof: (iii) and (iv) are obviously equivalent.

(ii) implies (i): Suppose (ii). Then, using Lemma 5.1, with $e_i = b_i$ and $f = \&(b\Rightarrow a)_J$, i in J, there is a P such that $P(\&(b\Rightarrow a)_J) = 1$ and $P(b_i) > 0$, i in J. This is sufficent, as similar reasoning in previous proofs show, to insure that (i) holds.

not(ii) implies not(i): Suppose not(ii). Then, there is an i in J such that $b_i\&(\&(b\Rightarrow a)_J) = \emptyset$. Hence, for any P so that $P(a_j|b_j) = 1$, all j in J, implies $P(\&(b\Rightarrow a)_J) = 1$, and thus implies by the above disjointness that $P(b_i) = 0$, a contradiction to $P(a_i|b_i) = 1$. Thus, not(i) must hold.

(iii) implies (ii): Obvious, since each $a_i \leq b_i$.

not(iii) implies not(i): Suppose not (iii). Then, there is some i in J with $a_i\&(\&(b\Rightarrow a)_J) = \emptyset$. Hence, if there is some P so that $P(a_j|b_j) =1$, all j in J, again this implies $P(\&(b\Rightarrow a)_J) = 1$, implying, in turn, from the above disjointness condition, that $P(a_i) = 0$, contradicting $P(a_i|b_i) = 1$. Thus, not(i) must hold. ■

Theorem 5.10. Characterization of SCPL deduction.

Under Basic Assumption I and SCPL consistency holding for $(a|b)_J$, the following statements are equivalent:

(i) $(a|b)_J \ \leq_{SCPL} (c|d)$.

(ii) $\underset{i \text{ in } J}{\mathrm{Or}}\big(\ \&(b\Rightarrow a)_J \leq \ (b_i \vee d)\Rightarrow c) \ \big)$.

Proof: not(i) holds iff (there is a P)$\big(\big($P$(a|b)_J = 1$) and (either P(d) = 0 or P(c|d) < 1)$\big)$
iff (there is a P)$\big($ (P$(\&(b\Rightarrow a)_J) = 1$) and (P$(b_J) > 0_J$) and
(either P(d′) = 1 or P(c′d) > 0) $\big)$
iff $\big($[(there is a P)$\big($ (P$(d'\&(\&(b\Rightarrow a)_J)) = 1$) and (P$(b_J) > 0_J$)$\big)$] or
[(there is a P)$\big($ (P$(\&(b\Rightarrow a)_J) = 1$) and (P$(b_J) > 0_J$) and (P(c′d) > 0) $\big)$]
iff, using Lemma 5.1 twice, with at first $f = d'\&(\&(b\Rightarrow a)_J)$ and $e_i =$ b_i, and then $f = \&(b\Rightarrow a)_J$, $e_i = b_i$, $e_o = c'd$, by extending index set J to include o corresponding to c′d, etc.,

$$[\ \underset{i \text{ in } J}{\mathrm{And}}\big(\ b_i d'\&(\&(b\Rightarrow a)_J) \neq \varnothing\big) \ ]$$

$$\text{or } \ [\ \underset{i \text{ in } J}{\mathrm{And}}\big(b_i\&(\&(b\Rightarrow a)_J) \neq \varnothing\big) \ \text{ and } \ \big(c'd\&(\&(b\Rightarrow a)_J) \neq \varnothing\big)]. \tag{5.35′}$$

Hence, the equivalence in eq.(5.35′) shows, by negating through,

$$\text{(i) holds iff } \ [\ \underset{i \text{ in } J}{\mathrm{Or}}\big(\ (\&(b\Rightarrow a)_J \leq \ b_i\Rightarrow d \ \big)$$
$$\text{and } [\ \underset{i \text{ in } J}{\mathrm{Or}}\big(\&(b\Rightarrow a)_J \ \leq \ b_i'\big) \ \text{ or } \ \big(\&(b\Rightarrow a)_J \ \leq \ d\Rightarrow c\big) \ ]$$
$$\text{iff } \ [\ \underset{i \text{ in } J}{\mathrm{Or}}\big(\&(b\Rightarrow a)_J \leq \ b_i'\big) \ \text{ or } \ \underset{i \text{ in } J}{\mathrm{Or}}\big(\&(b\Rightarrow a)_J \ \leq \ (b_i\vee d)\Rightarrow c\big) \ ]. \tag{5.36}$$

However, because of SCPL consistency (see Theorem 5.9(ii)) , we cannot have the left-hand side expression at the bottom of eq.(5.36), [Or$\big(\&(b\Rightarrow a)_J \leq \ b_i'\big)$], holding true there. Hence, (i) holds iff the bottom right-hand side expression of (5.36) holds, which is the desired result (ii). ■

Theorem 5.11. Basic relations among consistencies and deductions.

Under Basic Assumption I, the following hold:

(i) With respect to $(a|b)_J$: either SCPL consistency or SHPL consistency implies
WHPL consistency $=^w$ WCPL consistency.

(ii) Under also SHPL consistency for $(a|b)_J$,

$((a|b)_J \le_{WHPL} (c|d))$ iff $((a|b)_J \le_{SHPL} (c|d))$ implies $((a|b)_J \le_{WCPL} (c|d))$.

(iii) Under also WHPL consistency for $(a|b)_J$,

$((a|b)_J \le_{WHPL} (c|d))$ implies $((a|b)_J \le_{WCPL} (c|d))$.

Proof: Simply compare the algebraic forms in the previous Theorems of this section. ■

Theorem 5.12. Reduction of weak and strong CPL and HPL consistencies and deductions to the classical logic case for unconditional events.

Make Basic Assumption I, with the proviso that now $b_J = \Omega = d$. Then:

(i) $(a|\Omega)_J$ is SHPL consistent iff it is WHPL consistent iff it is WCPL consistent iff it is SCPL consistent iff $\&(a_J) \ne \emptyset$.

(ii) Under the common consistency assumption in (i),

$$((a|\Omega)_J \le_{SHPL} (c|\Omega)) \text{ iff } ((a|\Omega)_J \le_{WHPL} (c|\Omega)) \text{ iff } ((a|\Omega)_J \le_{WCPL} (c|\Omega))$$
$$\text{iff } ((a|\Omega)_J \le_{SCPL} (c|\Omega)) \text{ iff } \&(a_J) \le c,$$

the same as in classical logic (see, e.g., [Copi, 1986], where the *basic conjunctive deduction relation* is usually presented via the equivalent truth-table form of "whenever all a_j are verified (or true), then so must c (be true)".

Proof: Apply the simplifying constraint $b_J = \Omega = d$ to all of the previous relevant theorems in this section. ■

Theorem 5.13. The behavior of minconc when deduction fails in the HPL or CPL senses.

Under Basic Assumption I:

(i) Under also SHPL consistency for $(a|b)_J$: If not$((a|\Omega)_J \le_{SHPL} (c|\Omega))$, then not $(a|\Omega)_J \le_{SHPL} (c|\Omega))$, and for any real δ, $0 < \delta < 1$, slightly abusing notation by replacing $(1\text{-}\delta)\cdot 1_J$ by just $1\text{-}\delta$,

$$\text{minconc}_2\big((a|b)_J;(c|d)\big)(1\text{-}\delta) \le \delta.$$

(ii) Under WCPL consistency for $(a|b)_J$: If not $((a|\Omega)_J \le_{WCPL} (c|\Omega))$, then

$$\text{minconc}_2\big((a|b)_J;(c|d)\big)(1_J) = 0.$$

(iii) Under SCPL consistency, where also

$$\underset{i \text{ in } J}{\text{Or}}(\&(b\Rightarrow a)_J \le b_i \Rightarrow d),$$

then, [not $(((a|\Omega)_J \le_{SCPL} (c|\Omega))$] implies [$\text{minconc}_2\big((a|b)_J;(c|d)\big)(1_J) = 0$] .

Proof: Employ each algebraic characterization of the appropriate type of deduction (via the above theorems in this section) and consider the negation of each characterization and choose a P over this region, usually with the value 1. ■

As stated previously, in one way or another, Theorem 5.13 points out specifically the essentially extreme (0,1)-only possible values for the minconc function, for sufficiently high threshold values. Thus, the minconc function, despite its use as seen in the above theorems, *cannot* be interpreted as a way of measuring "degree of deduction" -- or softening of deduction of (c|d) from $(a|b)_J$ -- when, instead of taking limits of minconc as the premise thresholds approach unity, one holds them fixed and simply uses the evaluation of minconc as is. This leads to considering the meanconc function as a possibly more reasonable candidate to reflect such softening of deduction, as will be verified in the next section.

6. Some New Results and Insights in Expected Surety Logic

6.1 *Review of Relevant Definitions and Concepts Required*

Returning to the meanconc function and expected surety logic, consider the following: Suppose again we make Basic Assumption I, as given at the beginning of Section 5 and used throughout there. Recall also the discussion at the beginning of Section 2, where a set of m+1 atoms $\mathcal{A} = \{\alpha_1,...,\alpha_m,\alpha_{m+1}\} \subseteq \mathcal{B}$ (for given probability space $(\Omega,\mathcal{B},P)$) is given with respect to $((a|b)_J;(c|d))$, (and relative to all boolean combinations of the antecedents b_j and consequents a_j of each proper conditional event $(a_j|b_j)$, j in J). Unless otherwise stated, the designated atom $\alpha_{m+1} \leq \&(b'_J)$, therefore, assumed nonvacuous. For any probability measure P, we have its natural identification with the evaluations of P over atoms $\alpha_1,..., \alpha_m$, i.e., with values $(P(\alpha_1),...,P(\alpha_m)) = (x_1,...,x_m) =^d X^T$, with $P(\alpha_{m+1}) = x_{m+1} =$ 1-sum(X) – $(.)^T$ denoting vector or matrix transpose – with X lying in the m-simplex S_m of all possible values of such X: $0_m \leq X \leq 1_m$, sum(X) ≤ 1, so that as P varies, X varies, etc. Also, each event c in the boolean combinations generated by $\mathcal{A}$ can be uniquely expressed as a disjoint disjunction of certain of the atoms, written, $c = \vee(\alpha_{I(c)}) =^d \underset{j \text{ in } I(c)}{\vee}(\alpha_j)$ (using, again, multivariable notation) where index set $I(c) \subseteq \{1,..., m, m+1\}$ (with usually only the first m integers involved). When unambiguous, we will interchange P with X being in S_m, keeping in mind the last component x_{m+1} of P is not really in S_m. Also, as in the discussion in Section 2, we use the notation which identifies any relevant event as an m by 1 column vector of 0's and 1's with respect to the first m components of $\mathcal{A}$ (provided α_{m+1} is not part of it).

Next, for each real s, t, 0 < s, t < 1 and any real vector $\underline{s}$, of size card(J), define the following subsets of S_m:

$$\mathcal{A}_t =^d \{P \text{ in } S_m\text{: } P(a|b)_J \geq t\} =^w \{X \text{ in } S_m\text{: } h_1(X) \geq t\} \text{ (common lower bound t)}, \tag{6.1.1}$$

$$\mathcal{A}_{(\underline{s})} =^d \{P \text{ in } S_m\text{: } P(a|b)_J = \underline{s}\} =^w \{X \text{ in } S_m\text{: } h_1(X) = \underline{s}\}, \tag{6.1.2}$$

$$\mathcal{B}_t =^d \{P \text{ in } S_m\text{: } P_o(\&_o(a|b)_J) \geq t\} =^w \{X \text{ in } S_m\text{: } h_2(X) \geq t\}, \tag{6.1.3}$$

$$B_{(s)} =^{d} \{P \text{ in } S_m: P_o(\&_o(a|b)_J) = s\} =^{w} \{X \text{ in } S_m: h_2(X) = s\}, \tag{6.1.4}$$

$$C_t =^{d} \{P \text{ in } S_m: P_o(\&_{AC}(a|b)) \geq t\} =^{w} \{X \text{ in } S_m: h_3(X) \geq t\}, \tag{6.1.5}$$

$$C_{(s)} =^{d} \{P \text{ in } S_m: P_o(\&_{AC}(a|b)) = s\} =^{w} \{X \text{ in } S_m: h_3(X) = s\}. \tag{6.1.6}$$

Here,

$$h_1(X) =^{d} (h_{1,j}(X))_{j \text{ in } J} , \; h_{1,j}(X) =^{d} (a_j^T \cdot X / b_j^T \cdot X), \; j \text{ in } J, \tag{6.1.7}$$

are bounded bilinear functions in X. $h_2(X)$ as a function of X is obtained similarly as h_1, by replacing P everywhere appropriately by X in the computations in eq. (4.16′) and, while much more nonlinear in structure than $h_1(X)$, it is, nevertheless also a "well-behaved" function (i.e., differentiable, bounded, etc.). $h_3(X)$ is obtained likewise from eq.(4.27) with P replaced by X, noting that, just as $h_1(X)$, $h_3(X)$ is also a bounded bilinear function in X. Also, for any choice of prior (second order) probability distribution for X over S_m – such as typically here Dirichlet (including the uniform one over S_m as a special case) – denote the corresponding cdf's with respect to antecedent space A_t, B_t, C_t by $F_{i,t}$ and the corresponding cdf's with respect to $A_{(\underline{s})}$, $B_{(s)}$, $C_{(s)}$, by $G_{i,t}$, for any x in S_m (or, more generally, in m-dimensional real space),

$$F_{1,t}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } A_t) , \; G_{1,\underline{s}}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } A_{(\underline{s})}) , \tag{6.1.8}$$
$$F_{2,t}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } B_t) , \; G_{1,s}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } B_{(s)}) , \tag{6.1.9}$$
$$F_{3,t}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } C_t) , \; G_{1,s}(x) =^{d} \text{Prob}(X \leq x \mid X \text{ in } C_{(s)}). \tag{6.1.10}$$

6.2 *Method of Approach to Evaluation of Meanconc*

It is clear that for the fixed (i.e., non-unity limiting) threshold case the determination of the two types of meanconc functions – i.e., $E(P(c|d) \mid P \text{ in } A_t)$ or $E(P(c|d) \mid P \text{ in } A_{(t)})$ – requires evaluation in general of multiple integrals over the polytope formed from S_m and the constraints via A_t or $A_{(t)}$. Tables 1 and 2 illustrate closed-form results for a number of special types of premise sets and potential conclusions. Even there, in some cases, such as *transitivity*, full evaluation, under the simplest assumptions, such as choice of a uniformly distributed prior for P, required lengthy integration evaluations. (See Section 6.3 of [Goodman, 1999] for an outline of the calculations for the transitivity case.) On the other hand, there have been many advances in the area of such calculations, as seen, e.g., in [Bisztriczky *et al.*, 1994], which potentially can significantly reduce such calculations. (See also [Goodman & Nguyen, 2000] for further discussion.)

With the above in mind, let us consider an alternative to either attempting to obtain full closed-form evaluations or direct numerical approximations. The direction here is one of approximation, but in the following sense: We first attempt to show that, in an asymptotic sense, the expectation antecedents, (P in A_t), (P in $A_{(t)}$), are essentially equivalent in their effect upon the potential consequent P(c|d) as the related forms B_t, $B_{(t)}$ (where P satisfies $P_o(\&_o(a|b)_J) \geq t, = t$) and C_t, $C_{(t)}$ (where P satisfies $P(\&_{AC}(a|b)_J) \geq t, = t$). In addition, another justification for considering replacement of A_t (but not A(t), etc.) by the corresponding spaces determined by $\&_o$ or by $\&_{AC}$ is the "intertwining" – or asymptotic intertwining property they possess with respect to one another. That is, any one of the three spaces, A_t, B_t, C_t, determined by the separate

constraints $P(a_j|b_j) \geq t$, essentially lie inside any of the other two for t appropriately changed – but still retaining a value relatively close to unity. While this property alone does not guarantee convergence of the corresponding conditional expectations to each other as t approaches unity, it is an enhancing property of the closeness of the three spaces to each other asymptotically (see Theorem 6.5). Fortunately, Theorem 6.2 reinforces this closeness in demonstrating that, indeed, under mild conditions, all three spaces do lead asymptotically to the same relevant conditional distributions and expectations! While it is of some interest to know that $\&_o(a|b)_J$ could be used at least in theory for the desired replacement at some reasonably high (though not necessarily exactly unit) threshold level, inspection of its structure in eqs.(4.14)-(4.16′) compared to that of the far simpler – and, in fact, *single conditional event reducing* -- $\&_{AC}(a|b)_J$ (see eqs.(4.23)-(4.27)) shows the latter as the preferred candidate (even though it is not a true conjunction operator, etc.).

Ideally, one would then determine the replaced conditional expectation $E(P(c|d) \mid P(\&_{AC}(a|b) \geq t)$ (or $E(P(c|d) \mid P(\&_{AC}(a|b) = t))$. However, even in this case, preliminary results (at this point) indicate there is still considerable complexity of the required integration procedure – although these investigations do show a basic connection with the evaluation of certain integrals of hypergeometric functions of higher order. But, final success can be achieved in a modified way, by adapting an approach analogous to that employed by the popular naive maximum entropy approach $E(P(c|d) \mid P \text{ in } \mathcal{A}_t)$, where first a criterion is satisfied – i.e., maximizing the possible (first order probability) entropy and then the result is plugged into the objective function, i.e., the maximizing entropy probability measure P^* is then used to evaluate $P^*(c|d)$. Thus, as the counterpart to the above, we seek first to find that probability measure $P^\#$ which is most central to the premise set, i.e., we seek to obtain $P^\# = E(X| \text{ X in } \mathcal{A}_t)$ or the similar expression where $\mathcal{A}_t$ is replaced by $\mathcal{A}_{(t)}$ (see also the limiting forms in Theorem 6.2(iii) below) and then "plug-in" to evaluate $P^\#(c|d)$. For simplicity, we shall consider only $E(X|X \text{ in } A_{(t)})$ in the actual evaluations, carried out in Theorems 6.6 and 6.7.

The next result is stated in part in [Goodman & Nguyen, 2000] and is presented here in slightly different form for clarity. It is a generalization of a basic surface integral result due originally to [Higgins, 1975], with related work carried out by [Saw, 1973]:

Theorem 6.1. Decomposition into a weighted sum of ratios of surface integrals of any conditional cdf whose antecedent is generated from the implicit solution to a well behaved function being constant.

Let $(\Omega,\mathcal{B},P)$ be a real probability space, X an associated m by 1 random vector over some domain S in $(\text{Real line})^m$ with joint pdf f which is continuous and uniformly bounded over S. Let n be a positive integer with $n < m$ and let $h:S\to(\text{Real line})^n$ be a "well-behaved" function (uniformly bounded above and below away from zero, differentiable, etc.). Then, for any event $c \subseteq S$ and any real vector $\underline{s}$ in range(h),

$$P(X \text{ in } c \mid h(X) \leq \underline{s}) = \int_{\underline{r}\leq\underline{s}} \big(\rho(h,f;c)(\underline{r})\cdot g(h,f)(\underline{r})\big)d\underline{r} \quad \text{(ordinary integral)}, \tag{6.2.1}$$

where

$$\rho(h,f;c)(\underline{r}) \overset{d}{=} \psi(h,\underline{r};f,c)/\psi(h,\underline{r};f,S), \tag{6.2.2}$$

$$g(h,f)(\underline{r}) =^{d} \quad \psi(h,\underline{r};f,S) \,/ \int_{\underline{v}\le\underline{r}} (\,\psi(h,\underline{v};f,S)\,)d\underline{v}\ , \tag{6.2.3}$$

and surface integral

$$\psi(h,\underline{s};f,c) =^{d} \quad \int_{X \text{ in } surf_{h,s}\cap c} (f(x)\,/[\det(dh(X)/dX)(dh(X)/dX))^{T})])\ dsurf_{h,\underline{s}}(X)\ ; \tag{6.2.4}$$

$$surf_{h,\underline{s}} = \{X \text{ in } S\text{: } h(X) = \underline{s}\} = h^{-1}(\underline{s}). \tag{6.2.5}$$

(See, e.g., [Devinatz, 1968] for details of general surface integration; the surface integral in eq.(6.2.4) can also be converted back to ordinary integral form, but will not be needed here.) Note that the non-negative ("well-behaved") function in $\underline{r}$, $\rho(h,f;c)(\underline{r})$ is bounded by unity and that $g(h,f)(\underline{r})$ as a function of $\underline{r}$ is a legitimate pdf. ■

Remark.

Note first the vector derivative in eq.(6.2.4) is the n by m matrix of partial derivatives of the various scalar component function of h with respect to each single argument and, as part of the "well-behaved " property of h, it is assumed to be of full rank n (< m) so that the factor $\det(dh(X)/dX)(dh(X)/dX))^{T}) > 0$ in all X in $surf_{h,\underline{s}}$, for all $\underline{s}$ in range(h).

The original form of the above theorem does not appear as a weighted sum: the factor

$$\psi(h,\underline{r};f,S) = \int_{X \text{ in } surf_{h,s}\cap S} (f(x)\,/[\det(dh(X)/dX)(dh(X)/dX))^{T})])\ dsurf_{h,\underline{r}}(X)$$

$$= \int_{X \text{ in } surf_{h,s}} (f(x)\,/[\det(dh(X)/dX)(dh(X)/dX))^{T})])\ dsurf_{h,\underline{r}}(X) \tag{6.2.6}$$

cancels out in eq.(6.2.1).

Theorem 6.2. Three equivalent limiting forms involving meanconc.

Suppose that Basic Assumption I holds. For a second order probability prior over S_m, choose random vector X (representing a random probability measure) to be distributed so that its pdf over S_m is bounded and continuous. For example, we can choose the Dirichlet distribution dir($\underline{\tau}$) for X with parameter $\underline{\tau}$ such that $\underline{\tau} \ge 1$. Suppose also that

$$A_1 =^{d} \bigcap_{t\uparrow 1}(A_t) \ne \varnothing\ . \tag{6.2.7}$$

(The condition in eq.(6.2.7) can be analyzed via Theorem 5.9 for consistency of SHPL.) Then:

(i) Referring to eqs. (6.2.1)-(6.2.3), each of the three ordinary cdf's $F_{j,t}$ can be identified as weighted averages of the corresponding cdf's $G_{j,s}$ over the exact threshold spaces, all identified in the natural sense with the expansion in Theorem 6,1: For all x in S_m,

$$F_{1,t}(x) = \int_{\underline{t}\le \underline{s}\le 1} (G_{1,\underline{s}}(x).g_1(\underline{s}))d\underline{s}, \quad F_{2,t}(x) = \int_{t\le s\le 1} (G_{2,s}(x).g_2(s))ds, \quad F_{3,t}(x) = \int_{t\le s\le 1} (G_{3,s}(x).g_3(s))ds \ .$$

The above representations rigorize formal readily-derived counterparts which simply use integration-out of variables and chaining.

(ii) $\underset{\underline{t}\uparrow 1}{\text{limit}}\ (F_{1,\underline{t}}(x)) = G_{1,\underline{1}}(x)$, $\underset{t\uparrow 1}{\text{limit}}\ (F_{2,t}(x)) = G_{2,1}(x)$, $\underset{t\uparrow 1}{\text{limit}}\ (F_{3,t}(x)) = G_{3,1}(x)$,
where, for the limiting cdf's

$$G_{1,\underline{1}}(x) = G_{2,1}(x) = G_{3,1}(x) \ =^{d}\ G(x), \text{ all x in } S_m.$$

(iii) $\underset{t\uparrow 1}{\text{limit}}\ (E(X|\ X \text{ in } A_t)) = \underset{t\uparrow 1}{\text{limit}}\ (E(X|\ X \text{ in } B_t)) = \underset{t\uparrow 1}{\text{limit}}\ (E(X|\ X \text{ in } C_t))$
$= E(X)$ for X assigned cdf G.

(iv) $\underset{t\uparrow 1}{\text{limit}}\ (E(P(c|d)\ |\ P \text{ in } A_t)) = \underset{t\uparrow 1}{\text{limit}}\ (E(P(c|d)|\ P \text{ in } B_t)) = \underset{t\uparrow 1}{\text{limit}}\ (E(P(c|d)|\ P \text{ in } C_t))$

$= E(P(c|d))$ for P (or X) assigned cdf G.

Proof: First replace separately in Theorem 6.1, h by h_j, j=1, 2, 3. Also, replace there c by the infinite left ray at x in real m-space, and S by S_m. This yields (i). Then, noting that, because each cdf is a weighted sum of the corresponding exact threshold cdfs with range space $t \le s \le 1$, as a typical example, squeezing down to unity itself as $t \uparrow 1$, and that all cdf's are "well-behaved", etc., the top part of (ii) also holds. That the bottom part of (ii) holds is because, by inspection:

$h_1(X) = 1$ iff $P(a|b)_J = 1$; $h_2(X) = 1$ iff $P_o(\&_o(a|b)_J) = 1$ iff $P(a|b)_J = 1$;

$h_3(X) = 1$ iff $P_o(\&_{AC}(a|b)_J) = 1$ iff $P_o((\&_{AC}(a|b)_J)') = 0$ iff $P(\vee(a'b)_J) = 0$
iff $P(a_j'b_j) = 0$, all j in J (but $P(b_j) > 0$) iff $P(a|b)_J = 1$.

Finally, (iii) holds by simply applying the extended Helly-Bray moment theorem separately to each of the three converging sequences of cdf's for both the identity function in X and the bilinear function in X representing P(c|d). both being continuous bounded functions of X. (See, e.g., [Loève, 1963], Sections 11.3, 11.4.) ■

Theorem 6.2 insures the mutual asymptotic equivalence of meanconc with respect to either the original separate premise conditions, their PS conjunction forming one compound conditional form in PSCEA, or their *replacement by a significantly simpler single proper conditional via* $\&_{AC}$. However, the next results also reinforce this approximate equivalent asymptotic behavior in another direction.

Theorem 6.3. A lower bound on E(X in A_t | X in C_t).

Suppose Assumption I holds, as well as SCPL consistency (see Theorem 5.9). Suppose also that P, i.e., X, is distributed over S_m as Dirichlet dir($\underline{\tau}$), where parameter vector $\underline{\tau} = (\tau_1,\ldots,\tau_m,\tau_{m+1}) \geq 1_{m+1}$

Then, there exists an increasing function $g:[0,1]\to[0,1]$ with $g(0) = 0$, $g(1) = 1$, such that for all t, $0 < t < 1$,

$$g(t) \leq E_P\big(\mathrm{And}(P(a|b)_J \geq t) \mid P(\&_{AC}(a|b)_J) \geq t\big).$$

One can choose for g, without loss of generality (for at at least all t sufficiently close to unity)

$$g(t) = \big(1\big/\big(1+\sqrt{(1-t)/t}\big)\big)\cdot[1 - \max_{j \text{ in } J}\ (F_j\big(\sqrt{(1-t)/t}\,\big))], \tag{6.2.8}$$

where F_i is the cdf of the beta($\tau_{1,i}$,$\tau_{2,i}$) distribution;

$$\tau_{1,i} =^d \sum_{j \text{ in } I(\varphi_i)} (\tau_j)\,;\ \ \varphi_i =^d a_i\&(\&(b\Rightarrow a)_J) \neq \varnothing;\quad \tau_{2,i} =^d \sum_{j \text{ in } I(A\neg\varphi_i)} (\tau_j)\,. \tag{6.2.9}$$

Proof: Since $\&_{AC}(a|b) = (A \mid A \vee A'B)$, $A'B = \&(b')_J$, with A defined as in eq.(6.2.9′),

$$A =^d (\&(b\Rightarrow a)_J)\&(\vee(b)), \tag{6.2.9′}$$

$P((\&_{AC}(a|b)) \geq t$ iff $P(A'B) \leq ((1-t)/t)P(A)$. Hence,

$$P(a_i|b_i) \geq P(\varphi_i)/(P(\varphi_i)+ P(A'B)) \ \geq\ U_i / (U_i + ((1-t)/t))\,, \tag{6.2.10}$$

where, from the theory of Dirichlet distributions applied to random vector P (see, e.g., [Goodman & Nguyen, 1999a] or [Wilks, 1963]),

$$U_i =^d P(\varphi_i) / P(A) \text{ is distributed as beta}(t_{1,i}, t_{2,i})\text{, independent of } P(A) \text{ and } P(A'B). \tag{6.2.11}$$

Hence, denoting the pdf for beta($t_{1,i}$,$t_{2,i}$) as h_i, the required expectation here is

$$E(U_i / (U_i + ((1-t)/t))) = \int_{u=0}^{1} [\,\big(u/(u+((1-t)/t)\big)\cdot h_i(u)]du. \tag{6.2.12}$$

Breaking up the range of integration in (6.2.12) into two parts, [0, $\sqrt{(1-t)/t}$] and [$\sqrt{(1-t)/t}$, 1], it is clear that u/(u+((1-t)/t)) over the first interval varies between t and $\big(1\big/\big(1+\sqrt{(1-t)/t}\big)\big)$ while u/(u+((1-t)/t)) over the second interval varies from 0 to $\big(1\big/\big(1+\sqrt{(1-t)/t}\big)\big)$. Since $\underline{\tau} \geq 1_{m+1}$, there is a unique finite maximum m_i given as

$$m_i = h_i((\tau_{1,i} - 1)/(\ \tau_{1,i} - 1 + \tau_{2,i} - 1)),\ i \text{ in } J, \tag{6.2.13}$$

with limiting (uniform distribution) case interpreted as 1, when $\tau_{1,i} = \tau_{1,i} = 1$. (For justification of eq.(6.2.13), see [Johnson & Kotz, 1972], vol. 2, Chapter 24, Section 3.)

Finally, putting eqs.(6..2.10)-(6.2.13) together,

$$(1/(1+\sqrt{(1-t)/t}))\cdot(1-F_i(\sqrt{(1-t)/t})) \le E_P(\mathrm{And}(P(a|b)_J \ge t) \mid P(\&_{AC}(a|b)_J) \ge t), \quad \text{for i in J.} \qquad \blacksquare$$

Theorem 6.4. Some additional bounds connected with Theorem 6.3.

Under the same assumptions as in Theorem 6.3, for all $0 < t < 1$, for all i in J,

(i) $\mathrm{Var}(P(a_i|b_i) \mid P(\&_{AC}(a|b)_J) \ge t) \le (1-g(t))\cdot E(P(a_i|b_i) \mid P(\&_{AC}(a|b)_J) \ge t) \le 1-g(t)$.
(ii) For all real ε, $0 < \varepsilon < 1$,

$$\mathrm{Prob}\,(|P(a_i|b_i) - E(P(a_i|b_i) \mid P(\&_{AC}(a|b)_J)\ge t)| \le \varepsilon) \ge 1-(1-g(t))/\varepsilon^2.$$

(iii) $\mathrm{Prob}\big(P(a_i|b_i) \ge g(t)-(1-g(t))^{1/3} \mid P(\&_{AC}(a|b)_J) \ge t\big) \ge 1-(1-g(t))^{1/3}$,

for all t sufficiently close to 1.

(iv) $\mathrm{Prob}\big(\underset{\text{i in J}}{\mathrm{And}}[P(a_i|b_i) \ge g(t)-(1-g(t))^{1/3}] \mid P(\&_{AC}(a|b)_J) \ge t\big) \ge 1-(1-g(t))^{1/3}$,

for all t sufficiently close to 1.

Proof: Straightforward use of Theorem 6.3 together with Chebychev's inequality, where, letting $Y_i =^d P(a_i|b_i)$, $Z_{i,t} =^d E(Y_i \mid C_t)$, $\sigma^2_{i,t} =^d \mathrm{Var}(P(a_i|b_i) \mid P(\&_{AC}(a|b)_J) \ge t)$,

$$\begin{aligned}\mathrm{Prob}(Y_i \ge g(t) - \varepsilon \mid C_t) &\ge \mathrm{Prob}(Y_i \ge Z_{i,t}-\varepsilon \mid C_t) \ge \mathrm{Prob}(|Y_i-Z_{i,t}| \le \varepsilon \mid C_t) \ge 1-\sigma^2_{i,t}/\varepsilon^2 \\ &\ge 1-(1-g(t))/\varepsilon^2\end{aligned}$$

and then choosing $\varepsilon = (1-g(t))^{1/3}$. $\blacksquare$

Definition. Given any probability space (Ω, B, P) and two collections $A = (A_t)_{0<t<1}$, $B = (B_t)_{0<t<1}$ with A_t, B_t in B and each collection nesting down as t increases. Then, say that A and B are *intertwined* iff there are functions f, g: $[0,1]\to[0,1]$ increasing and continuous with $f(0) = g(0) = 0$, $f(1) = g(1) = 1$, such that

$$B_{f(t)} \le A_t \le B_{g(t)}, \text{ for all t, } 0 < t < 1.$$

Hence, conversely,

$$A_{g^{-1}(t)} \le B_t \le A_{f^{-1}(t)}, \text{ for all t, } 0 < t < 1.$$

Lemma 6.1 Let $A = (A_t)_{0<t<1}$, $B = (B_t)_{0<t<1}$ be two intertwining collections as in the definition. Let

$A_1 =^d \underset{0<t<1}{\&}(A_t)$ and $B_1 =^d \underset{0<t<1}{\&}(B_t)$. Then, $A_1 = B_1$. Hence, if one is nonvacuous, so will be the other.

Proof: $A_1 = \underset{0<t<1}{\&}(A_t) \le \underset{0<t<1}{\&}(B_{g(t)}) = \underset{0<t<1}{\&}(B_t) = B_1 \le \underset{0<t<1}{\&}(A_{f^{-1}(t)}) = \underset{0<t<1}{\&}(A_t) = A_1$. ■

Theorem 6.5. Intertwining and almost intertwining of A_t's, B_t's, C_t's as given in eqs.(6.1), (6.3), (6.5).

Since the A_t's, B_t's, C_t's are sets of probability vectors (the P's or the X's in our notation), we use ordinary subset notation here. Under the same assumptions of Theorem 6.3, the following relations hold for any t, 0 < t < 1,

$$A_t \subseteq B_{1-(1-t)\mathrm{card}(J)} \subseteq C_{1-(1-t)\mathrm{card}(J)}, \quad A_{1-(1-t)/\mathrm{card}(J)} \subseteq B_t \subseteq C_t, \tag{6.2.14}$$

noting that C_t's in eq.(6.2.14) dominate all subset inclusions. As a partial converse, where g(t) is provided from Theorem 6.3,

$$\mathrm{Prob}\,(X \text{ in } A_{g(t)-(1-g(t))^{1/3}} | \, X \text{ in } C_t) \ge 1 - (1-g(t))^{1/3},$$

i.e.,

$$\mathrm{Prob}(\, C_t \subseteq A_{g(t)-(1-g(t))^{1/3}}) \ge 1 - (1-g(t))^{1/3}. \tag{6.2.15}$$

Proof: Straightforward use of FHH inequalities (see eqs. (1.1.5) and (5.22)) and Theorem 6.4. ■

Theorem 6.6. Closed-form expression for E(P| P(a|b) = t)

Make the Basic Assumption I, where now J = {1}, $A_{(t)}$ = {P: P(a|b) = t}, and vector partition X as $X^T = (x_1,...,x_m) = (X_{(1)}^T, X_{(2)}^T, X_{(3)}^T)$, $X_{(1)}^T = (x_1,...,x_n)$, $X_{(2)}^T = (x_{n+1},...,x_{n+p})$, $X_{(3)}^T = (x_{n+p+1},...,x_m)$, where, as usual $x_j = P(\alpha_j)$, x_{m+1} = 1- sum(X), etc. Suppose also that X is distributed over S_m as Dirichlet dir($\underline{\tau}$), where $\underline{\tau} = (t_1,...,t_m,t_{m+1}) > 0_{m+1}$. Then, E(P| P(a|b) = t) = E(X | X in $A_{(t)}$) is in $A_{(t)}$ and

$$E(X|\, X \text{ in } A_{(t)})^T = E(P \mid P(a|b) = t)^T = (\, E(X_{(1)} \mid A_{(t)})^T, E(\, X_{(2)} \mid A_{(t)})^T, E(X_{(3)} \mid A_{(t)})^T\,),$$

where

$$E(X_{(1)} \mid A_{(t)})^T = t\cdot w_{(1)}\cdot(w_1,...w_n)\, , E(X_{(2)} \mid A_{(t)})^T = (1-t)\cdot w_{(1)}\cdot(w_{n+1},...w_{n+p})\, ,$$
$$E(X_{(3)} \mid A_{(t)})^T = w_{(2)}\cdot(w_{n+p+1},...,w_m)\, , E(x_{m+1} \mid A(t)) = 1-\mathrm{sum}(E(X|A_{(t)})) = 1-w_{(1)}-w_{(2)} = w_{(3)},$$

where

$$w_{(1)} =^d (\tau_{(1)} + \tau_{(2)}) / (\tau_{(1)} + \tau_{(2)} + \tau_{(3)} + \tau_{m+1})\, , w_{(2)} = \tau_{(3)} / (\tau_{(1)} + \tau_{(2)} + \tau_{(3)} + \tau_{m+1}),$$
$$w_{(3)} = \tau_{m+1} / (\tau_{(1)} + \tau_{(2)} + \tau_{(3)} + \tau_{m+1}),$$
$$\tau_{(1)} =^d \tau_1+...+\tau_n, \quad \tau_{(2)} =^d \tau_{n+1}+...+\tau_{n+p}, \quad \tau_{(3)} =^d \tau_{n+p+1}+...+\tau_m$$
$$w_1 =^d (1/\tau_{(1)})\cdot\tau_1,..., w_n =^d (1/\tau_{(1)})\cdot\tau_n\, ; \quad w_{n+1} =^d (1/\tau_{(2)})\cdot\tau_{n+1},..., w_{n+p} =^d (1/\tau_{(2)})\cdot\tau_{n+p},$$
$$w_{n+p+1} =^d (1/\tau_{(3)})\cdot\tau_{n+p+1},..., w_m =^d (1/\tau_{(3)})\cdot\tau_m..$$

Proof: Since $A_{(t)}$ is a (closed) convex set, E(X | X in $A_{(t)}$) is also in $A_{(t)}$. Next, consider three cases: x_i for i = 1,...,m, letting U = sum($X_{(1)}$), V = sum($X_{(2)}$), W = sum($X_{(3)}$), and in this notation, $A_{(t)}$ holds iff U/(U + V) = t:

Case 1. i = 1,..., n.

$E(x_i|A_{(t)}) = E\big((x_i/U)\cdot(U/(U+V))\cdot(U+V) \mid U/(U+V) = t\big)$.

But, from the theory of Dirichlet distributions (see, again, [Wilks, 1963] or [Goodman & Nguyen, 1999a]), the random variables (x_i/U), U/(U+V), (U+V) are all independent of each other and have the beta distribution. Since the middle one is determined from the antecedent of the expectation to be t, we need only specify the first, which is
beta($\tau_i,\tau_{(1)}$-t_i) and the third which is beta($\tau_{(1)}+\tau_{(2)},\tau_{(3)}+\tau_{m+1}$), so that

$$E(x_i|A_{(t)}) = (\tau_i/\tau_{(1)})\cdot t\cdot((\tau_{(1)}+\tau_{(2)})/(\tau_{(1)}+\tau_{(2)}+\tau_{(3)}+\tau_{m+1})).$$

Case 2. i=n+1,...,n+p.

$$E(x_i|A_{(t)}) = E\big((x_i/V)\cdot(V/(U+V))\cdot(U+V) \mid U/(U+V) = t\big)$$
$$= E\big((x_i/V)\cdot(1-(U/(U+V)))\cdot(U+V) \mid U/(U+V) = t\big),$$

noting the mutual independence and beta distributions of (x_i/V), U/(U+V), (U+V), with the middle term determined from the antecedent of the expectation as 1-t. Again, as in case 1, the beta distribution expectations are readily obtained.

Case 3. i=n+p+1,...,m

$E(x_i|A_{(t)}) = E\big((x_i/W)\cdot W \mid U/(U+V) = t\big)$,

noting the mutual independence and computable beta distributions of (x_i/W), W, U/(U+V). ■

Theorem 6.7. Plug-in evaluation $P^{\#}(c|d)$ relative to $P^{\#} =^{d} E(P \mid P(a|b) = t)$.

Under the same hypothesis as Theorem 6.6, noting again $P^{\#}$ = E(P | P(a|b) = t) is a legitimate probability vector in $A_{(t)}$ because of the closed convexity of $A_{(t)}$ = {P: P(a|b) = t}. Using the multivariable notation $\Sigma(W)_{I(c)}$ for $\sum_{j \text{ in } I(c)}(w_j)$, where I(c) ⊆ {1,...,m+1} is the index set for c with respect to set of atoms A_o,

$P^{\#}(c|d) = N^{\#} / D^{\#}$;

$N^{\#} = P^{\#}(c) = P^{\#}(ac) + P^{\#}(a'bc) + P^{\#}(b'c)$

$= t \cdot w_{(1)} \cdot \Sigma(w)_{I(ac)} + (1\text{-}t) \cdot w_{(1)} \cdot \Sigma(w)_{I(a'bc)} + w_{(2)} \cdot \Sigma(w)_{I(b'c)}$,

$D^{\#} = P^{\#}(d) = P^{\#}(c) + P^{\#}(c'd) = N^{\#} + P(ac'd) + P^{\#}(a'bc'd) + P^{\#}(b'c'd)$
$= N^{\#} + t \cdot w_{(1)} \cdot \Sigma(w)_{I(ac'd)} + (1\text{-}t) \cdot w_{(1)} \cdot \Sigma(w)_{I(a'bc'd)} + w_{(2)} \cdot \Sigma(w)_{I(b'c'd)}$

Proof: Straightforward substitution of values from Theorem 6.6 into $P^{\#}(c|d)$, using the decomposition $P^{\#}(c|d) = P^{\#}(c) / (P^{\#}(c) + P^{\#}(c'd))$, $P^{\#}(c) = (P^{\#}(ac) + P^{\#}(a'bc) + P^{\#}(b'c))$, $P^{\#}(c'd) = P^{\#}(ac'd) + P^{\#}(a'bc'd) + P^{\#}(b'c'd)$. ■

7. Summary of Algebraic Characterization of Asymptotic Form of Expected Surety Deduction for Common Threshold Case

Apropos to the discussion in Section 6.2 concerning the difficulty in exactly determining meanconc functions for the fixed nonlimiting threshold case, quite a different story holds for the evaluation of these functions when the thresholds are allowed to approach unity. [Bamber, 2000], under a uniform second order prior distribution assumption, in the same spirit of this paper, has derived an algebraic characterization for deciding when $E(P(c|d) \mid P \text{ in } \mathcal{A}_t)$ approaches unity as threshold t approaches unity. This algebraic procedure (involving in Bamber's terminology, "rarity functions") is equivalent to (also, algebraic) deduction-validating procedures in Pearl's [1990] *System Z* and in Lehmann & Magidor's [1992] *rational closure*. (In fact all of the above work for common threshold t has been extended to non-identical thresholds approaching unity via some power of one another – see, e.g., [Bamber, 2000] for further details.) We describe below, in outline form, an equivalent process, which also allows us to obtain the full asymptotic distributional form associated with meanconc. This process, is a sequential one, where in step one given potential EPL conclusion (c|d) is first tested as to which combination of $c\&(\&(b\Rightarrow a)_J)$ and/or $c'd\&(\&(b\Rightarrow a)_J)$ is null or not and then, only if the first level indeterminate case holds, i.e., $c\&(\&(b\Rightarrow a)_J) = c'd\&(\&(b\Rightarrow a)_J) = \varnothing$, then one proceeds to refine this further by replacing J in step one by $K_1 = \{i \text{ in } J\text{: } a_i \,\&((\&(b\Rightarrow a)_J) = \varnothing\}$ (involving the SCPL criterion for the K_1 index set – see again, Theorem 5.9). Then, the procedure is repeated with again one continuing on to the next level only if the indeterminate case $c\&(\&(b\Rightarrow a)_{K_1}) = c'd\&(\&(b\Rightarrow a)_{K_1})$ $= \varnothing$. Re-examining the above a little more closely and making the usual general Dirichlet second order prior probability assumption for P (or X), one can readily obtain the full asymptotic limiting distribution of $(P(c|d) \mid \mathcal{A}_t)$ as t approaches unity: Returning to the first stage, consider

Step 1, Case 1. $c\&(\&(b\Rightarrow a)_J) \neq \varnothing$ and $c'd\&(\&(b\Rightarrow a)_J) \neq \varnothing$. By simply decomposing c and c′d relative to P(c|d) into their intersections with $\&(b\Rightarrow a)_J)$ and its complement $\vee(a'b)_J)$ and noting that the constraints in $\mathcal{A}_t$ show that for t close to unity any probability assigned to anything intersecting some $a_j'b_j$ will be negligible (unless all probabilities are so). This shows that a typical P(c|d) is essentially the same as

$$P(c\&(\&(b\Rightarrow a)_J)) / [P(c\&(\&(b\Rightarrow a)_J)) + P(c'd\&(\&(b\Rightarrow a)_J))]$$

which has no further constraints upon it as t approaches unity and, from basic Dirichlet family properties, as a random variable has a beta distribution with nontrivial computable parameters.

Hence, in general, for this case neither unity nor zero mass point asymptotic distributions occur in general.

Step 1, Case 2. $c\&(\&(b\Rightarrow a)_J) = \varnothing$, i.e., $c \leq (\vee(a'b)_J)$, and $c'd\&(\&(b\Rightarrow a)_J) \neq \varnothing$.

By reasoning similar to Case 1, the asymptotic form of a typical P(c|d) must be

$$P(c\&(\vee(a'b)_J) / [P(c\&(\vee(a'b)_J) + P(c'd\&(\vee(a'b)_J)],$$

where clearly $P(c\&(\vee(a'b)_J)$ approaches zero, but $P(c'd\&(\vee(a'b)_J)]$ remains independent of t. Hence, this case stochastically leads to a zero mass-point for the limiting form of $(P(c|d) \mid A_t)$.

Step 1, Case 3. $c\&(\&(b\Rightarrow a)_J) \neq \varnothing$, and $c'd\&(\&(b\Rightarrow a)_J) = \varnothing$. i.e., $c'd \leq \vee(a'b)_J$.

The asymptotic limiting form here for $(P(c|d)|A_t)$ is

$$P(c\&(\&(b\Rightarrow a)_J) / P(c\&(\&(b\Rightarrow a)_J),$$

i.e., a unity mass-point distribution.

Step 1, Case 4. $c\&(\&(b\Rightarrow a)_J) = c'd\&(\&(b\Rightarrow a)_J) = \varnothing$.

Replacing J above in the four cases of Step 1 by K_1, leads to the following possibilities:

Step 2, Case 1. $c\&(\&(b\Rightarrow a)_{K_1}) \neq \varnothing$, and $c'd\&(\&(b\Rightarrow a)_{K_1}) \neq \varnothing$.
yet,

$$c\&(\&(b\Rightarrow a)_J) = \varnothing, \text{ and } c'd\&(\&(b\Rightarrow a)_J) = \varnothing.$$

Then, analogous to the reasoning for Step 1, Case 1, while both numerator and denominator of P(c|d) go to zero as t approaches unity, the *rate of convergence* to zero for $P(c\&(\&(b\Rightarrow a)_{K_1})$ and $P(c'd\&(\&(b\Rightarrow a)_{K_1})$ remain an order of magnitude less than the complement terms. Whence, the same formal situation once more holds in that the dominating expression for P(c|d) is

$$P(c\&(\&(b\Rightarrow a)_{K_1}) / [P(c\&(\&(b\Rightarrow a)_{K_1}) + P(c'd\&(\&(b\Rightarrow a)_{K_1})]$$

which is beta distributed asymptotically, etc.

One then continues, just as in each Case of Step 1, establishing the lower rate of zero convergent expressions. If Case 4 of Step 2 holds, then we must refine even further, replacing now K_1 by $K_2 = \{i \text{ in } K_1\text{: } a_i \&((\&(b\Rightarrow a)_{K_1}) = \varnothing\}$ and continuing the process which is guaranteed to end in a finite number of steps.

8. Closing Remarks and Some Research Issues.

Analogous to Section 5, where various relations were established between the weak and strong forms of HPL and CPL deduction, one can determine various relations between EPL deduction and HPL and CPL deduction. At the outset, as in Theorem 5.12 for CPL and HPL, EPL deduction reduces to classical logic deduction relative to premise and potential conclusion consisting of only unconditional events. For simplicity, we have tacitly only considered the strong form of EPL, where all conditional probabilities involved are well-defined and will continue to consider that case here for HPL and CPL, as well, when possible. First, it is clear from the very definition of minconc and meanconc (see Section 3) that we always have

$$\mathrm{minconc}((a|b)_J;(c|d)(t) \leq \mathrm{meanconc}((a|b)_J;(c|d)(t), \quad \text{for } 0< t \leq 1. \tag{8.1}$$

Eq.(8.1) immediately implies that

$$(a|b)\leq_{HPL} (c|d) \quad \text{implies} \quad (a|b) \leq_{EPL} (c|d), \tag{8.2}$$

where the converse does not hold in general as seen in Table 1, with transitivity, contraposition, and strengthening being examples of EPL deductions but not HPL ones. On the other hand, Theorem 6.2(iv) shows that

$$(a|b)\leq_{EPL} (c|d) \quad \text{implies} \quad (a|b) \leq_{CPL} (c|d). \tag{8.3}$$

As in eq.(8.2), the implication in eq.(8.3) is not reversible in general. A case where CPL (weak) deduction is valid, but EPL is not, is provided by the *Nixon diamond* scheme (number 23 in Table 1). (Bamber [2000] has also considered relations between EPL, HPL, and CPL deduction validity.)

We illustrate here the nonmonotonicity of EPL vs. the monotonicity of HPL and CPL, where again, it should be noted that, by their very definitions, once a valid deduction holds in the sense of HPL or CPL, with respect to P(c|d) for given $(a|b)_J$, it must hold in the same sense for any given increased premise set $(a|b)_{J\cup K}$. Consider the *Penguin Triangle Deduction Scheme* (number 17 of Table 1), where in our general deduction notation (c|d) is (a′b |c), J = {1,2,3,4}, with $(a_1|b_1)$ being (a|b), $(a_2|b_2)$ being (b|c), $(a_3|b_3)$ being (d|c), and $(a_4|b_4)$ being (a′b |d), with as in all of Table 1, a, b, c, d *not* otherwise constrained (as opposed to Assumption I) First, form $\&(b\Rightarrow a)_J = b'c'd' \vee abc'd'$. Then, as in the procedure in Section 7, test for which case holds in Step 1 for $\&(b\Rightarrow a)_J$ conjoining here, nonvacuously or not, a′c, ac. Thus, only Step 1, Case 4 holds where both intersections are null. Then, form $K_1 = \{i \text{ in } J: a_i \,\&\, (b'c'd' \vee abc'd') = \varnothing\} = \{2,3,4\}$. In turn, now form $\&(b\Rightarrow a)_{K_1} = c'd \vee a'bd$ and retest for which possible nonvacuous-vacuous combination with a′bc, abc holds. This yields only Step 2, Case 3, since $a'bc\&c'd\,(c'd \vee a'bd) = a'bcd \neq \varnothing$, $abc\&(\,d \vee a'bd) = \varnothing$. Hence, in the limit as t approaches 1, $(P(c|d)|A_t)$ approaches the mass-point 1, i.e., $E(P(c|d)|A_t)$ approaches 1 and EPL validity holds, where A_t corresponds to J={1,2,3,4} here.

On the other hand, for the same potential conclusion (a′b |c), but smaller premise set than above, consisting only of that part of A_t corresponding to J = {1,2}, we know from the *transitivity* property (number 13, Table 1), that (a|c) is EPL-deduced from {(a|b), (b|c)} and hence $\underset{t\uparrow 1}{\mathrm{limit}}\big(E(P(a|c) \mid P(a|b), P(c|d) \geq t)\big) = 1$ and thus $\underset{t\uparrow 1}{\mathrm{limit}}\big(E(P(a'b \mid c) \mid P(a|b), P(c|d) \geq t)\big) = 0$, showing invalidity for the same conclusion with the smaller premise class.

Finally, among the key open problems, mention should be made of that of determining bounds on the differences between the actual meanconc functions and their asymptotic single rule-plug-in replacements. In a similar direction, upper bounds are sought in the general case – for both fixed thresholds and unity limiting ones -- for the differences between meanconc and alternative deduction functions, including minconc and maxent. While some of the discrepancies for certain specific cases among the three approaches (using meanconc, minconc, maxent) are pointed out in Tables 1 and 2, a more general study is needed to consider tradeoffs between desirable deduction properties and computational complexity. It is also of interest to consider extensions of EPL to a linguistic setting, as a counterpart to the numerous fuzzy logic approaches to reasoning. An outline for the beginning of such an extension is provided in Section 3.3 of [Goodman & Nguyen, 1999b].

9. References.

[Adams, 1966] E.W. Adams, "Probability and the logic of conditionals", in *Aspects of Inductive Logic* (J. Hintikka & P. Suppes, eds.), North-Holland, Amsterdam.

[Adams, 1975] E.W. Adams, *The Logic of Conditionals*, D. Reidel, Dordrecht, Holland, 1975.

[Adams, 1986] E.W. Adams, "On the logic of high probability", *Journal of Philosophical Logic*, vol. 15, pp. 255-279.

[Adams, 1996] E.W. Adams, "Four probability-preserving properties of inferences", *Journal of Philosophical Logic*, vol. 25, 1996, pp. 1-24.

[Aitchison, 1986] J. Aitchison, *The Statistical Analysis of Compositional Data*, Chapman & Hall, London, 1986.

[Alefeld & Herzberger, 1983] G. Alefeld & J. Herzberger, *Introduction to Interval Computations*, Academic Press, New York.

[Bamber, 2000] D. Bamber, "Entailment with near surety of scaled assertions of high conditional probability", *Journal of Philosophical Logic*, vol. 29, pp. 1-74, 2000.

[Bamber *et al.*, 2000] D. Bamber, I.R. Goodman & H.T. Nguyen, "Extension of the concept of propositional deduction from classical logic to probability: Part 1, an overview of probability-selection approaches", to appear in special publication based upon plenary sessions at Fifth Joint Conference on Information Sciences, Atlantic City, NJ, Feb. 27- March 3, 2000.

[Bisztriczky *et al.*, 1994] T. Bisztriczky, P. McMullen, R. Schneider & A.I. Weiss, *Polytopes: Abstract, Convex and Computational*, NATO ASI Series, Kluwer, Dordrecht, Holland.

[Buchanan & Shortliffe, 1984] B.G. Buchanan & E.H. Shortliffe, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, MA, 1984.

[Calabrese, 1987] P.G. Calabrese, "An algebraic synthesis of the foundations of logic and probability", *Information Sciences*, vol. 42, 1987, pp. 187-237.

[Calabrese, 1994] P.G. Calabrese, "A theory of conditional information with applications", *IEEE Trans. Systems Man & Cybernetics*, vol. 24(12), Dec., 1994, pp. 1676-1684.

[Copi, 1986] I.M. Copi, *Introduction to Logic, Seventh Edition*, Macmillan, New York
[Devinatz, 1968] A. Devinatz, *Advanced Calculus*, Holt, Rinehart & Winston, New York.
[Dempster, 1967] A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", Annals of mathematical Statistics, vol. 38, pp. 325-338.
[Goodman, 1999] I.R. Goodman, "A decision-aid for nodes in Command & Control Systems based on cognitive probability logic", Proceedings 1999 Command & Control Research & Technology Symposium, vol. 2 (held June 29 – July 1, 1999, Naval War College, Newport, RI), pp. 898-941.
[Goodman *et al.*, 1997] I.R. Goodman, R.P. Mahler & H.T. Nguyen, *Mathematics of Data Fusion*, Kluwer, Dordrecht, Holland
[Goodman & Nguyen, 1985] I.R. Goodman & H.T. Nguyen, *Uncertainty Models for Knowledge-Based Systems*, North-Holland, Amsterdam.
[Goodman & Nguyen, 1995] I.R. Goodman & H.T. Nguyen, "Mathematical foundations of conditionals and their probabilistic assignments", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, vol. 3(3), Sept. 1995, pp. 247-339.
[Goodman & Nguyen, 1998] I.R. Goodman & H.T. Nguyen, "Adams' high probability deduction and combination of information in the context of product probability conditional event algebra", *Proceedings International Conf. on Multisource-Multisensor Information Fusion (Fusion '98)*, Las Vegas, NV, July 6-9, 2000, vol.1, pp. 1-8.
[Goodman & Nguyen, 1999a] I.R. Goodman & H.T. Nguyen, "Probability updating using second order probabilities and conditional event algebra", *Information Sciences* vol.121, 1999, pp. 295-347.
[Goodman & Nguyen, 1999b] I.R. Goodman & H.T. Nguyen, "Application of conditional and relational event algebra to the defining of fuzzy logic concepts", *Signal Processing, Sensor Fusion, and Target Recognition VIII* (I. Kadar, ed.), *Proc. SPIE, vol. 3720*, April 5-7, 1999, Orlando, FL, pp. 25-36.
[Goodman & Nguyen, 2000] I.R. Goodman & H.T. Nguyen, "Computational aspects of quantitative second order probability logic and fuzzy if-then rules: part 1, basic representations as integrals", *Proc. Fifth Joint Conf. Info. Sci. (JCIS 2000)*, vol. 1, Feb. 27 – March 3, 2000, pp. 64-67.
[Goutsias *et al.*, 1997] J. Goutsias, R.P. Mahler & H.T. Nguyen, eds., *Random Sets: Theory & Applications* (I.M.A. series vol. 97), Springer, New York.
[Grove & Halpern, 1997] A.J. Grove & J.Y. Halpern, "Probability update: conditioning vs. cross-entropy", *Proceedings 13th Conference on Uncertainty in A.I.*, 1997, pp. 208-214.
[Hayes-Roth *et al.*, 1983] F. Hayes-Roth, D.A. Waterman, D.B. Lenat, *Building Expert Systems*, Addison-Wesley, Reading, MA.
[Hailperin, 1965] T. Hailperin, "Best possible inequalities for the probability of a logical function of events", *American Mathematical Monthly*, vol. 72, 1965, pp. 343-359.
[Hailperin, 1984] T. Hailperin, "Probability logic", *Notre Dame Journal of Formal Logic*, vol. 25(3), July 1984, pp. 198-212.
[Hailperin, 1996] T. Hailperin, *Sentential Probability Logic*, Lehigh University Press, Bethlehem, PA.
[Higgins, 1975] J.J. Higgins, "A geometrical method of constructing multivariate densities and some related inferential procedures", *Communications in Statistics*, vol. 4(10), pp. 955-966.

[Johnson & Kotz, 1972] N. Johnson & S. Kotz, *Distributions in Statistics: Vol. 2, Continuous Univariate Distributions, Part 2; 4, Continuous Multivariate Distributions*, Wiley, New York, 1973.
[Kapur, 1994] J.N. Kapur, *Measures of Information and Their Applications*, Wiley, New York.
[Lehmann & Magidor, 1992] D. Lehmann & M. Magidor, "What does a conditional knowledge base entail?", *Artificial Intelligence*, vol. 55, pp. 1-60.
[Lewis, 1976] D. Lewis, "Probabilities of conditionals and conditional probabilities", *Philosophical Review*, vol. 85, pp. 297-315.
[Loève, 1963] M. Loève, *Probability Theory, Third. Ed.*, Van Nostrand, Princeton, NJ.
[Mendelson, 1970] E. Mendelson, *Boolean Algebra and Switching Circuits, Schaum's Outline Series, McGraw-Hill*, New York, 1970.
[Nguyen, 1978] H.T. Nguyen, "On random sets and belief functions*", Journal of Mathematical Analysis & Its Applications*, vol. 65, pp. 531-542.
[Pearl, 1988] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo, CA, 1988.
[Pearl, 1990] J. Pearl, System Z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning", in *Theoretical Aspects of Reasoning about Knowledge, Proceedings of the Third Conference (TARK, 1990)* (R. Parikh, ed.), Morgan Kaufmann, San Mateo, CA, pp. 121-135.
[Prior *et al.*, 1967] N. Prior *et al.*, "History of logic", *in Encyclopedia of Philosophy, Vol. 4* (P. Edwards, *et al.*, eds.), Macmillan, New York, pp. 513-571.
[Rao, 1973] C.R. Rao, *Linear Statistical Inference & Its Applications, 2nd Ed.*, Wiley, New York, 1973.
[Rödder, 2000] W. Rödder, "Conditional logic and the principle of entropy", Artificial Intelligence, vol. 117, 2000, pp. 83-106.
[Saw, 1973] J.G. Saw, "Jacobians of singular transformations with applications to statistical distribution theory", *Communications in Statistics*, vol. 1(1), pp. 81-91.
[Schlechta, 1997] K. Schlechta, *Nonmonotonic Reasoning: Basic Concepts, Results and Techniques*, Springer, New York.
[Shafer, 1976] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.
[Van Fraasen, 1981] B.C. Van Fraasen, "A problem for relative information minimizers in probability kinematics", *British Journal for Philosophy of Science*, vol. 32, 375-379.
[Walley, 1991] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London, U.K.
[Wilks, 1963] S. Wilks, *Mathematical Statistics*, Wiley, New York, 1963.

Appendix A. Proof of Theorem 5.6.

Lemma A.1. Make Basic Assumption I, assume WHPL consistency, and for some K, $\emptyset \neq K \subseteq J$, assume that

$$\&(b\Rightarrow a)_K \&(\vee(b)_K) = \emptyset. \tag{A.1}$$

Then, for any real δ, $0 < \delta < 1/(2\mathrm{card}(J))$, and any P (probability measure over *B*), the following statements are equivalent:

(i) For each j in J, either $P(a_j|b_j) \geq 1-\delta$ or $P(b_j) = 0$.
(ii) For all j in J, $P(b_j) = 0$.

Proof: Obviously, (ii) implies (i). Suppose (i) holds for some P such that there is a set L, $\emptyset \neq L \subset K$ (proper) such that $P(a_j|b_j) \geq 1-\delta$, for all j in L and $P(b_j) = 0$, for all j in $K\neg L$ ($\neq \emptyset$). By the FHH lower bound (eq.(5.22)), we must have for PSCEA P_o counterpart of P,

$$\begin{aligned} P_o(\&_o(a|b)_L) &\geq \max(\Sigma(P(a|b)_L) - (\mathrm{card}(L)-1), 0) \geq \max(\mathrm{card}(L)\cdot(1-\delta) - (\mathrm{card}(L)-1), 0) \\ &= \max(1-\delta\cdot\mathrm{card}(L), 0) = 1-\delta\cdot\mathrm{card}(L) > ½ > 0. \end{aligned} \tag{A.2}$$

Then, eq.(A.2) combined with eq.(4.26) and the monotonicity property of probability shows that

$$P_o(\&_{AC}(a|b)_L) \geq P_o(\&_o(a|b)_L) > 0. \tag{A.3}$$

From the structure of $\&_{AC}$ in eqs.(4.23) and (4.27) combined with eq.(A.3),

$$P(\&(b\Rightarrow a)_L \&(\vee(b)_L) > 0. \tag{A.4}$$

But, the assumption $P(b_j) = 0$, j in $K\neg L$, implies $P(\&(b'_{K\neg L})) = 1$, whence using eq.(A.4),

$$P[\&(b\Rightarrow a)_L \&(\vee(b)_L)\&(\&(b'_{K\neg L}))] > 0. \tag{A.5}$$

Now, from the definitions in eqs. (4.16)-(4.18),

$$\begin{aligned} &\&(b\Rightarrow a)_L \&(\vee(b)_L)\&(\&(b'_L)) = \bigvee_{(\emptyset\neq C\subseteq L)} \left(\gamma(a,b;C,L) \,\&\, (b')_{K\neg L}\right) \\ &= \bigvee_{(\emptyset\neq C\subseteq L)} \left(\gamma(a,b;C,K)\right) \leq \bigvee_{(\emptyset\neq C\subseteq K)} \left(\gamma(a,b;C,K)\right) = \&(b\Rightarrow a)_K \&(\vee(b)_K). \end{aligned} \tag{A.6}$$

Thus, applying the monotonicity of probability to eq.(A.6) and combining with eq.(A.5) shows

$$P(\&(b\Rightarrow a)_K \&(\vee(b)_K)) > 0,$$

contradicting the assumption that $\&(b\Rightarrow a)_K \&(\vee(b)_K)) = \emptyset$. Hence, no such L can exist.
The only remaining possibilities are L = K or L = $\emptyset$. In the case of L = K, again application of the FHH lower bounds and again the ordering between $\&_{AC}$ and $\&_o$ in eq.(4.26) shows

$$P_o(\&_{AC}(a|b)_K) \geq P_o(\&_o(a|b)_K) > 0\ ,$$
which, analogous to eqs.(A.3), (A.4) implies

$$P(\&(b\Rightarrow a)_K \& (\vee(b)_K)) > 0,$$

once more, leading to a contradiction of the assumption. Hence, the only possibility is for (ii) to hold. ■

Lemma A.2. Under the same assumptions of Lemma A.1 and for the K satisfying eq.(A.1), assume

$$\&(b\Rightarrow a)_K \leq d\Rightarrow c. \tag{A.7}$$

Then, for any real δ, $0 < \delta < 1/(2\mathrm{card}(J))$, and any P (probability measure over *B*):

$$\text{If [for each j in J, either } P(a_j|b_j) \geq 1\text{-}\delta \text{ or } P(b_j) = 0\text{], then } [P(d\Rightarrow c) = 1]. \tag{A.8}$$

Proof: The "if" part of eq.(A.8), certainly implies the condition holds for K as a subset of J. Then, applying Lemma A.1, we must have $P(b_j) = 0$, all j in K, i.e.,

$$P(\&(b')_K) = 1. \tag{A.9}$$

But, combining eq.(A.7) with a standard property of the material conditional,

$$\&(b')_K \leq \&(b\Rightarrow a)_K \leq d\Rightarrow c. \tag{A.10}$$

Then, applying P throughout eq.(A.10) and using eq.(A.9) shows the desired result. ■

Lemma A.3. Make Basic Assumption I, assume WHPL consistency, and for some K, $\varnothing \neq K \subseteq J$, assume that

$$\varnothing \neq \&(b\Rightarrow a)_K \& (\vee(b)_K) \leq\ c \tag{A.11}$$

and

$$\&(b\Rightarrow a)_K \leq d\Rightarrow c. \tag{A.12}$$

Then, for that K,

(i) $\varnothing_o \neq \&_{AC}(a|b)_K \leq_o (c|d)$.

(ii) For any real δ, $0 < \delta < 1/(2\mathrm{card}(J))$ and any P such that for all j in J, either $P(a_j|b_j) \geq 1\text{-}\delta$ or $P(b_j) = 0$, then $[P(c|d) \geq 1 - \delta{\cdot}\mathrm{card}(J)]$.

Proof: First, the left-hand side of eq.(A.11) shows that $\&_{AC}(a|b)_K$ must be a proper conditional event. (Again, see the discussion following eq.(4.23).) Using the definition of $\&_{AC}$ and comparing the remaining part of eq.(A.11) and eq.(A.12) with the basic ordering criterion of PSCEA between proper conditional events given in eq.(4.19) shows the validity of (i).
Next, let $K_P =^d \{j \text{ in } K: P(b_j) = 0\}$. Thus, by hypothesis, $K\neg K_P = \{j \text{ in } K: P(a_j|b_j) \geq 1\text{-}\delta\}$,

$$P(\vee(b)_K) = P(\vee(b)_{K\neg K_p}), \quad P(\&(b\Rightarrow a)_K\&(\vee(b)_K)) = P(\&(b\Rightarrow a)_{K\neg K_p}\&(\vee(b)_{K\neg K_p})),$$

implying

$$P(\&_{AC}(a|b)_K) = P(\&_{AC}(a|b)_{K\neg K_p}). \tag{A.13}$$

Case 1. $K_P = K$. Then, by hypothesis (A.11),

$P(c) \geq P(\&(b\Rightarrow a)_K\&(\vee(b)_K) = P(\&(b')_K) = 1$, implying $P(c|d) = 1$.

Case 2. $\varnothing \subseteq K_P \subset K$ (proper).

Subcase 1. $P(\&(b\Rightarrow a)_{K\neg K_p}\&(\vee(b)_{K\neg K_p})) = 0$. But, by Lemma A.1, this condition implies for all j in $K\neg K_P$ that $P(b_j) = 0$, contradicting the very meaning of $K\neg K_P$.

Subcase 2. $P(\&(b\Rightarrow a)_{K\neg K_p}\&(\vee(b)_{K\neg K_p})) > 0$. Now, again using the FHH lower bound in eq.(5.22), as in eq.(A.2), replacing L there by $K\neg K_P$, combining with the monotonicity property of P_o applied to result (i) and the order relation between $\&_o$ and $\&_{AC}$, and using eq.(A.13),

$$P(c|d) \geq P_o(\&_{AC}(a|b)_K) = P_o(\&_{AC}(a|b)_{K\neg K_p}) \geq P_o(\&_o(a|b)_{K\neg K_p}) \geq 1 - \delta\cdot\mathrm{card}(K\neg K_p) \geq 1 - \delta\cdot\mathrm{card}(J),$$

the desired result for (ii). ■

Lemma A.4. Make Basic Assumption I and assume $(a|b)_J$ is WHPL consistent. Then,

$$\text{Assumption Q implies } [\, (a|b)_J \leq_{WHPL} (c|d)\,],$$

where

Assumption Q :

$$(\text{there is some } K, \varnothing \neq K \subseteq J)\big(\, [\&(b\Rightarrow a)_K\&(\vee(b)_K) \leq c\,] \text{ and } [\&(b\Rightarrow a)_K \leq d\Rightarrow c]\,\big).$$

Proof: Break up Assumption Q into two parts

$$Q_1: (\text{there is some } K, \varnothing \neq K \subseteq J)\big(\, [\&(b\Rightarrow a)_K\&(\vee(b)_K) = \varnothing\,] \text{ and } [\&(b\Rightarrow a)_K \leq d\Rightarrow c]\,\big),$$

$$Q_2: (\text{there is some } K, \varnothing \neq K \subseteq J)\big(\, [\varnothing \neq \&(b\Rightarrow a)_K\&(\vee(b)_K) \leq c] \text{ and } [\&(b\Rightarrow a)_K \leq d\Rightarrow c]\,\big),$$

and apply Lemma A.2 to Q_1 and Lemma A.3 to Q_2. ■

Lemma A.5. Make Basic Assumption I and assume $(a|b)_J$ is WHPL consistent. Then,

$$\text{not}(Q) \text{ implies } \text{not}[(a|b)_J \leq_{WHPL} (c|d)].$$

Proof: First note that

$$\text{not}(Q) \quad \text{iff} \quad (\text{for all } K, \varnothing \neq K \subseteq J)\big(\, I_{K,1} \text{ or } I_{K,2} \text{ or } I_{K,3}\big), \tag{A.14}$$

where

$$I_{K,1} =^d (\tau_K\&c'd \neq \varnothing)\,;\; I_{K,2} =^d (\&(b')_K\&c'd \neq \varnothing)\,;\quad I_{K,3} =^d (\tau_K\&d' \neq \varnothing); \tag{A.15}$$

$$\tau_K =^d \&(b\Rightarrow a)_K\&(\vee(b)_K), \qquad \varnothing \neq K \subseteq J. \tag{A.16}$$

Next, proceed to construct mutually disjoint "blocks", analogous to the procedure in the proof of Theorem 5.1 ((ii) implies (iv)), but slightly modified, beginning with index set J. Thus, we obtain a nonvacuous exhaustive disjoint partitioning $\{K_1,...,K_M\}$ of J, for some positive integer M, with the same notation as in eq.(5.15), such that there is a collection of mutually disjoint nonvacuous events – which are the blocks - given as

$$\gamma(a,b;Kj,J\neg K(j))\&\eta_j\ ,\ j = 1,...,M\text{-}1, \tag{A.17}$$

and where for the first time, at step M, by definition, either,

Case 1 $\gamma(a,b;K_M,J\neg K(_M))\&\eta_M = \&(a_{K_M})\&\eta_M$ is also nonvacuous and mutually disjoint with respect to events in (A.17),

where

$$\eta_j \text{ in } \{c'd,\ d'\},\ \ j = 1,...,M; \tag{A.18}$$

or

Case 2 $\gamma(a,b;\emptyset,J\neg K(M\text{-}1)) = \&(b')_{J\neg K(M-1)}$ is also nonvacuous and mutually disjoint with respect to events in (A.17), where all $\gamma(a,b;L,K)$ are defined as usual as in eq.(4.16) with K replaced by L and J by K, etc.

The procedure ends when either Case 1 or Case 2 first occurs. (In the construction in the proof of Theorem 5.1, the conjunctive factors η_j were missing and the procedure ended when Case 1 first occurred –at the Mth step.) Next, analogous to the construction of P in the proof of Theorem 5.1 ((iv) implies (I)), assign

$$\begin{aligned}
&P[\gamma(a,b;Kj,J\neg K(j))\&\eta_j] =^d \delta^{j-1} - \delta^j,\ \text{for } j = 1,...,M\text{-}1;\\
&P[\gamma(a,b;K_M,J\neg K(M\text{-}1))\&\eta_M] =^d \delta^M,\ \text{if Case 1 holds;}\\
&P[\gamma(a,b;\emptyset,J\neg K(M\text{-}1))] =^d \delta^M,\ \text{if Case 2 holds.}
\end{aligned} \tag{A.19}$$

Then, analogous to the proof of Theorem 5.1 (see eqs.(5.19)-(5.21)), it follows that

$$P(a_j|b_j) \geq 1\text{-}\delta,\ \text{for all j in J.} \tag{A.20}$$

In addition, since only either $c'd$ or d' (but not both) appears explicitly at each block_j, j=1,...,M, one has only the possibilities:

Situation 1: There is at least some block j in which $c'd$ appears. Then, clearly, since P is unity over the disjunction of all of the blocks,

$$P(c'|d) = P(c'd)/P(d) \geq \sum_{(1\leq i\leq M,\ c'd \text{ appears at block } i)} P(\text{block}_i) \Big/ \Big(\sum_{(1\leq i\leq M,\ c'd \text{ appears at block } i)} P(\text{block}_i) \Big) = 1. \tag{A.21}$$

Hence, eqs.(A.20) and (A.21) show that not$[(a|b)_J \leq_{WHPL} (c|d)]$, due to the arbitariness of δ.

Situation 2. There is no block in which $c'd$ appears, i.e., only d' appears in each of the blocks.

Thus, in this situation,

$$\varnothing \neq \tau_{K_i}\&d', \; i = 1,...,M, \tag{A.22}$$

But, eq.(A.22) implies

$$\varnothing \neq \tau_{K_i}, \; i = 1,...,M. \tag{A.23}$$

Then, according to Theorem 5.1, eq(A.23) insures that $(a|b)_J$ is SHPL consistent. On the other hand, eq.(A.22) and the construction of P in eq.(A.19), shows, analgous to eq.(A.21) that

$$P(d') = 1. \tag{A.24}$$

Also, eqs.(A.20) and (A.24) imply that not$[(a|b)_J \leq_{SHPL} (c|d)]$ (we need $P(c|d)$ positive and "high" for SHPL to hold). In turn, because of SHPL consistency, Theorem 5.4 shows that not$[(a|b)_J \leq_{SHPL} (c|d)]$ implies not$[(a|b) \leq_{WHPL} (c|d)]$. Hence, finally, not$[(a|b) \leq_{WHPL} (c|d)]$ also holds in this situation. ■

Finally:

Proof of Theorem 5.6: Simply combine Lemmas A.4 and A.5. ■